Possibility and Impossibility of Learning with Limited Behavior Rules

Takako Fujiwara-Greve*
Dept. of Economics
Keio University, Tokyo, Japan and
Norwegian School of Management BI, Sandvika, Norway
and
Carsten Krabbe Nielsen
Instituto de Politica Economica
Universita Cattolica, Milan, Italy.

Preliminary version: February, 2004.

Abstract: We consider boundedly rational learning processes in which players have a priori limited set of behavior rules. A behavior rule is a function from information to a stage-game action, which reflects the available information and one's reasoning about how others act. Commonly used behavior rules include the adaptive rule and the conservative rule (inertia). Sophisticated players may use iterative best responses, called forward-looking behavior rules. The feasible behavior rules set the framework and limitations of learning processes.

We investigate a general relationship between the set of feasible behavior rules and the properties of the long-run outcomes of any learning process restricted to use the feasible behavior rules. From very limited set of behavior rules to increasingly sophisticated ones, robust limit actions are what we call minimal weak-curb sets. In order to converge to a minimal weak-curb set of arbitrary stage game, it is sufficient that the players have one period memory and use one-step different behavior rules. For some classes of games where minimal weak-curb sets coincide with Nash equilibria, we have a global convergence to a Nash equilibrium under very limited set of behavior rules. For other games, however, there is a limit to learning. We show that for any finite set of behavior rules and finite memory length, there is a class of stage games whose unique Nash equilibrium cannot be reached from some initial states. More sophistication of reasoning does not mean better convergence in the sense that the process stays in a smaller set or finds a minimal curb set, which is a set-generalization of Nash equilibrium. The key to finding curb sets is not only the depth of thinking but also the stage game payoff structure. That is, smartness alone cannot find a rational action without the aid of a payoff incentive to explore actions.

JEL classification number: C73.
Key words: learning, game, adaptive, forward-looking, curb set.

*Corresponding author. Email: takako.greve@bi.no. Mailing address: Elias Smiths vei 15, Box 580, Sandvika, N-1302, NORWAY. Fax: + 47 - 6755-7675. (Until September, 2004.)

1 Introduction

We look at learning processes of boundedly rational players with multiple ways of reasoning about how others act. Players with multiple ways of reasoning ("behavior rules") not only fit human learning better[1] but also may learn better than players with a fixed algorithm such as the adaptive process. In fact, Hurkens [5] showed that if players can use infinitely-iterative best responses, they can eventually play a minimal weak-curb set, which is a set-generalization of Nash equilibrium. It is, however, also plausible that players can only reason finitely many steps due to bounded rationality. Therefore we take the middle ground between a single algorithm and the infinitely-iterative reasoning.

To see why an interest in a set of behavior rules playing against a (possibly different) set of behavior rules matters, consider the design of collision avoidance systems in airplanes. These systems use the radar to predict a potential collision, such as with an approaching airplane, and then proceed to warn the pilot or take evasive action on their own. The evasive action might be fixed; such as making a controlled descent, but the obvious problem is that an oncoming plane might have the same system and might also make a fixed descent. Or the system may randomly pick between an ascent and a descent. Or one plane may not have a collision avoidance system, as in most small planes. Or the pilot of one or both of the planes may decide to override the collision avoidance system based on his or her inference about what the other plane might do. In order to avoid collisions in a variety of potential encounters, the designer needs to choose between different types of information to react to and different levels of intelligence to build in. It is not obvious that more intelligence is better – indeed a collision avoidance system replaces an intelligent actor (a pilot) with a less intelligent one (an algorithm).

Our main focus is on belief-based behavior rules. A belief-based behavior rule prescribes an action which is a best response to some belief generated by the available information. A notable example is the adaptive behavior rule, which maps a best response to one of the observed actions by the opponent. There are other belief-based behavior rules. For example, a (one-step) forward-looking behavior rule maps a best response to the expected action by an opponent

[1]Roth et al. [11] and Selten [13] give experimental evidence that human subjects act diversely in the same game. Stahl [15] and [16] give both theoretical and experimental analysis of rule learning, in which players have multiple ways of reasoning and may change reasoning over time.

using the adaptive rule. One can iterate many steps of forward-looking reasoning.[2] We consider classes of learning processes using various combinations of these behavior rules. Many known learning processes are special cases of our model.[3] To isolate the effect of behavior rules on the long-run outcomes, we exclude random actions ("experimentation" in actions) which many models assume.[4]

From very limited set of behavior rules to increasingly sophisticated ones, robust limit actions are what we call minimal weak-curb sets: minimal product sets which are closed under best responses to pure actions. A singleton weak-curb set is a strict Nash equilibrium. In order to converge to a minimal weak-curb set of arbitrary stage game, it is sufficient that the players have one period memory and use diverse behavior rules (e.g., conservative and adaptive behavior rules, adaptive and one-step forward-looking behavior rules) with positive probabilities. In particular, if the stage game has only singleton minimal weak-curb sets[5], we have a global convergence to a strict Nash equilibrium using only simple behavior rules.

By contrast, minimal curb sets, which are closed under best responses to all *mixed* actions, are not easily reached under a limited set of behavior rules. We show that for any finite set of behavior rules and finite memory length, there is a class of stage games whose unique Nash equilibrium (a minimal curb set) cannot be reached because the process is stuck in a different minimal weak-curb set. A minimal curb set is a generalization of a Nash equilibrium, while a minimal weak-curb set may not contain the support of any Nash equilibrium. Hence in general it is difficult to learn to play a Nash equilibrium action under limited sets of behavior rules.

A remedy is to introduce even more sophisticated behavior rules as Hurkens [5] did, but the

[2]Another line of reasoning is backwards. A conservative behavior rule prescribes one of the observed actions of your own population. A performance-based behavior rule prescribes the observed action with the highest payoff. In the main model, we focus on one-period memory and thus backward-looking behavior rules become a degenerate conservative rule. See also the extension in Section 4.

[3]For example, finite memory best response dynamics (such as Cournot dynamic and Young's no mistake process [18]) are processes in which the behavior rule is restricted to be the adaptive one. Josephson [6] considers a class of learning processes with slightly different behavior rules from ours, consisting of imitators, adaptive players, and better repliers. In his model the shares of behavior rules are fixed over time. Matros [8] considers adaptive processes with clever agents, in which a fixed share of one of the populations use one-step forward-looking rule and the rest use the adaptive behavior rule. A common idea among these models is that a player has a fixed behavior rule over time. By contrast, we allow players to keep or alter the behavior rules over time.

[4]See Marimon and McGrattan [7] for a survey.

[5]For example coordination games, weakly acyclic games (Young [18]) and finite-action supermodular games (Milgrom and Roberts [9]) have this property. See Section 3.2.

necessary level of sophistication (in terms of the number of iterations in best responses and the range of possible behavior rules of the opponent that one allows in his beliefs) depends on the stage game. Adding more sophistication and/or behavior rules increases the volatility of the process, but it may cycle in a larger set than a minimal weak-curb set and still away from Nash equilibria. (See the example of Table 8 in Section 3.4.) Therefore, more sophistication does not always mean better convergence in the sense that the process stays in a smaller set or finds minimal curb sets. The obstacle to learning is not only the limited ability in reasoning but also the payoff structure of the game. That is, smartness alone cannot find a rational action without the aid of a payoff incentive to explore actions.

Finally, we note how our results complement the literature. Young [18] (the no-mistake model) showed that when the adaptive process has long enough[6] memory as compared to sampling, it converges to a Nash equilibrium in weakly acyclic games, in which minimal weak-curb sets coincide with Nash equilibria. We considered general games and insufficient memory and added diversity in behavior rules. Hurkens [5] (the sophisticated learning model) showed convergence to a minimal curb set when any level of forward-looking behavior rule is feasible. We showed that, under subsets of those behavior rules, the convergence does not hold. Milgrom and Roberts [10] showed that rationalizable strategies are long-run outcomes of processes consistent with sophisticated learning. With slightly more structure imposed on the processes, we have a sharper prediction that the actions converge to a minimal weak-curb set, which is smaller than the set of rationalizable strategies. Weak-curb sets are *self-confirming* in the class of iterative best-response behavior rules with one-period memory, since (i) the actions are best response to the belief and (ii) the beliefs are consistent with the (limited) behavior rules and the information. Therefore it is a natural limit, just like Fudenberg and Levine [3] argue for learning of extensive-form games.

The paper is organized as follows. In Section 2, we show motivating examples to illustrate the ideas of convergence and non-convergence. In Section 3, we give formal results of one-period memory model. In Section 4, we extend the results to general finite-memory model.

[6]The sufficient condition depends on the stage game.

2 Motivating Examples

Example 1: 2 by 2 Coordination games.

There are two populations, Pop 1 and Pop 2, from which one player is randomly chosen to play the role of player 1 and player 2 of a 2 by 2 coordination game (see Table 1 below) in periods $t = 1, 2, \ldots$. The payoffs satisfy $a_1, a_2, b_1, b_2 > 0$.

P1 \ P2	a	b
A	a_1, a_2	$0, 0$
B	$0, 0$	b_1, b_2

Table 1: 2 by 2 (normalized) Coordination games

If the players use the *adaptive* behavior rule throughout the time and they have one-period memory, the action profile may cycle between (A, b) and (B, a). The adaptive behavior rule prescribes a best response to the previous period action by the opponent.

Consider the *one-step forward-looking* behavior rule, which prescribes a best response to the opponent who uses the adaptive behavior rule. For example, if (A, b) is observed, the adaptive behavior rule for Pop 1 player specifies action B (the best response to action b by the Pop 2 in the previous period), while the one-step forward-looking behavior rule for Pop 1 player specifies action A (the best response to action a by the Pop 2 player who uses the adaptive behavior rule). Hence, if both players use one-step forward-looking rule throughout the time, the action profile may cycle, just like the case that both use the adaptive rule throughout.

If there is a positive probability that one player uses the adaptive rule and the other uses one-step forward-looking rule, then the action profile can move to one of the Nash equilibria, from any non-Nash states (A, b) and (B, a). For example, if the observation was (A, b), then it is sufficient that player 1 uses the adaptive rule to play action B and player 2 uses one-step forward-looking rule to play action b, or player 2 uses the adaptive rule to play action a and player 1 uses one-step forward-looking rule to play action A.

Moreover, if one player uses the *conservative behavior rule,* which chooses the same action as the previous period, and the other player uses the adaptive behavior rule, then again the action profile can converge to one of the pure Nash equilibria.

The convergence is thanks to the "coordination" of reactions by the different behavior rules. Note also that it was not the sophistication of forward-looking behavior rule that helped the convergence since the conservative rule can also yield convergence, but the diversity in behavior rules contributed to the convergence.

Example 2: Idea of Global Convergence

Using the following stage game as an example, we describe the idea of global convergence from any initial observation. Assume that players have one-period memory and only use an iterative-best-response behavior rule or the conservative behavior rule. In addition, assume that there is a probability $\epsilon > 0$ that, in each period, one of the players uses the adaptive behavior rule and at the same time the other player uses the conservative behavior rule.[7] This ϵ is time and state independent.

P1 \ P2	a	b	c	d
A	$2,2$	$0,1$	$0,1$	$0,1$
B	$1,0$	$3,3$	$1,0$	$1,0$
C	$1,0$	$1,0$	$0,2$	$2,1$
D	$1,0$	$2,0$	$0,1$	$1,2$

Table 2

There are 16 possible observations or *states* . We divide them into 3 classes:

• $\{(A,a),(B,b)\}$. Once one of the strict Nash equilibria is observed, any iterative best-response leads to the same action and thus the process converges.

• $[\{B,C,D\}\times\{a\}]\cup[\{A,C,D\}\times\{b\}]\cup[\{A\}\times\{b,c,d\}]\cup[\{B\}\times\{a,c,d\}]$. (One of the observed actions is a strict Nash equilibrium action.) If a player who observed the strict Nash equilibrium action by the opponent uses the adaptive behavior rule and the other player uses conservative behavior rule, then the resulting action profile is a strict Nash equilibrium, i.e., the process converges to one of the strict Nash equilibria in one period, with probability at least ϵ. For

[7] Analogously, we can use one-step forward-looking behavior rule instead of the conservative rule.

example, if (A, c) is observed, Pop 2 player uses the adaptive behavior rule, and Pop 1 player uses the conservative behavior rule, then the next period action profile is (A, a).

In the coordination game Example 1, only these two classes of states exist.

• $\{C, D\} \times \{c, d\}$. (None of the observed actions belong to a strict Nash equilibrium.) In this case we may need more than one step to reach a strict Nash equilibrium.

First, notice that $\{C, D\} \times \{c, d\}$ is not closed under best response, that is, there is an action c such that if it is observed, the best response to it is outside of $\{C, D\}$. If such "exit" action is observed, it is possible that player 1 using the adaptive rule chooses action B and player 2 using the conservative behavior rule chooses action c. Then the process enters the above (second) class of the states in one period, and hence it reaches one of the strict Nash equilibria in two periods. This occurs with probability at least ϵ^2.

Second, consider the subset $\{C, D\} \times \{d\}$. This is a Cartesian product of actions and is not closed under best response. In particular, if C is observed, player 2 using the adaptive behavior rule would play c, which is outside of $\{d\}$. If (C, d) is observed, then it is possible that player 2 uses the adaptive rule and plays c, while player 1 uses the conservative behavior rule to play C. In this case the process reaches a strict Nash equilibrium (B, b) in three periods. If (D, d) is observed, then it is possible that player 1 uses the adaptive rule to play action C and player 2 uses the conservative behavior rule to play action d. From there, the process can reach (B, b) in three periods. The move $(C, d) \to (C, c) \to (B, c) \to (B, b)$ or $(D, d) \to (C, d) \to (C, c) \to (B, c) \to (B, b)$ occurs with probability at least ϵ^4.

This logic can be iterated if there are more (but finite) actions in this class.

In sum, there is a (time and state independent) positive probability $p = \epsilon^4 > 0$ such that, from any state, the process reaches a strict Nash equilibrium within four periods. In other words, the probability that the process does not reach one of the strict Nash equilibria in $4t$ periods is at most $(1 - p)^t$, which converges to zero as t tends to infinity, i.e., the process converges to one of the strict Nash equilibria almost surely.

Example 3: A game with non-degenerate minimal curb set.

The above two examples have a special feature that the minimal curb sets (Basu and Weibull [1]) coincide with strict Nash equilibria. The following game has a minimal curb set which is not a strict Nash equilibrium. This game has two Nash equilibria; a strict Nash equilibrium (C, c) and a mixed Nash equilibrium $((1/2)A * (1/2)B, (1/2)a * (1/2)b)$.

P1 \ P2	a	b	c
A	$4, 0$	$0, 4$	$0, 1$
B	$0, 4$	$4, 0$	$0, 1$
C	$1, 0$	$1, 0$	$1, 1$

Table 3: A game with non-degenerate minimal curb set

The product set of actions $\{A, B\} \times \{a, b\}$ is closed under rational behavior (Basu and Weibull [1]). For any belief with the support $\{a, b\}$, player 1's best response is contained in $\{A, B\}$. Similarly, for any belief with the support $\{A, B\}$, player 2's best response is contained in $\{a, b\}$. Moreover, there is no smaller product set within $\{A, B\} \times \{a, b\}$ which has this property. Hence $\{A, B\} \times \{a, b\}$ is a minimal curb set. Not coincidentally, $\{A, B\} \times \{a, b\}$ corresponds to the support of the mixed Nash equilibrium. As Hurkens [5] discusses, minimal curb sets is a set-valued generalization of strict Nash equilibria. Clearly, the strict Nash equilibrium (C, c) is a singleton minimal curb set.

If players use only the adaptive behavior rule throughout the time, the action profiles may cycle among $\{(A, c), (C, b), (B, c), (C, a)\}$ and never reaches any of the minimal curb sets. If players use the adaptive behavior rule and the conservative behavior rule with positive probabilities, then it is possible to reach a minimal curb set from any initial information, just like the previous example. Moreover, once a process enters a minimal curb set, it will not leave the set as long as the players use a best-response based behavior rule or the conservative rule. (The iterated best responses are all contained in the minimal curb set.)

Example 4: A game with a minimal weak-curb set, which is not a curb set.

Consider a slightly different game in Table 4. This game has a unique and strict Nash equilibrium (C, c). The set $\{A, B\} \times \{a, b\}$ is no longer a curb set since some beliefs have a best response outside of it. However, it is closed under best response to itself, i.e., the best responses to a or b are contained in $\{A, B\}$ and the best response to A or B are contained in $\{a, b\}$. We call such set a weak-curb set (see Section 3). A strict Nash equilibrium is a (singleton) minimal weak-curb set as well as a minimal curb set.

P1 \ P2	a	b	c
A	$4, 0$	$0, 4$	$0, 0$
B	$0, 4$	$4, 0$	$0, 0$
C	$3, 0$	$3, 0$	$1, 1$

Table 4: A game with a minimal weak-curb set, which is not a curb set

By the same logic as in the previous examples, if the players use the adaptive and conservative behavior rules with positive probabilities, a process reaches one of the minimal weak-curb sets. Moreover, if the players use only the behavior rules that prescribe an iterated best response to the observation or the conservative rule, then a process never leaves a minimal weak-curb set even if it is not curb. This is because these behavior rules put probability one on one of the actions in $\{A, B\} \times \{a, b\}$ as the belief.

Therefore, even if players can use sophisticated behavior rules which compute many-step iterated best responses, a minimal weak-curb set cannot be left. There is a limit to learning with singleton-belief formation and best response dynamics. There is no payoff incentive to play action C as long as players use this type of simple iterative reasoning.

Now consider a *mixed* behavior rule, which incorporates a probability distribution of various behavior rules that the opponent may use. For example, Pop 1 player can believe that the opponent uses one of the conservative and the adaptive behavior rules but not for sure. If the probability $w_1 \in [0, 1]$ that the opponent is expected to use the adaptive rule is specified, this reasoning generates a best response to a mixed action. (This type of behavior rule is called an *M2 rule* with weight parameter w_1.)

If (A, a) is observed and player 1 uses M2 rule with $w_1 = 1/2$, then (s)he plays a best response to $(1/2)a * (1/2)b$, which is action C. Therefore the action profile can get out of the non-curb set $\{A, B\} \times \{a, b\}$ under some mixed behavior rule. This suggests that if players can enlarge the set of behavior rules in the direction of more actions being rationalized, it may be possible to get out of non-curb sets.

Example 5: Impossibility of finding a minimal curb set.

Although for a simple game like the one in Table 4, it was sufficient to introduce M2 rules, one can infer that the necessary behavior rules to warrant convergence to minimal curb sets is stage-game dependent. To put it differently, if we fix the highest level of iteration of best responses, there is a class of games in which no learning process can reach the unique Nash equilibrium (minimal curb set) from some initial observation.

Consider M3 rules such that a player has a belief over the conservative behavior rule and M2 rules by the opponent. The game in Table 5 has a unique and strict Nash equilibrium (singleton minimal curb set) (E, e). The set $\{A, B, C, D\} \times \{a, b, c, d\}$ is a minimal weak-curb set but not curb.

P1 \ P2	a	b	c	d	e
A	$7, 0$	$0, 0$	$0, 0$	$0, 1$	$-1, -1$
B	$0, 1$	$7, 0$	$0, 0$	$0, 0$	$-1, -1$
C	$0, 0$	$0, 1$	$7, 0$	$0, 0$	$-1, -1$
D	$0, 0$	$0, 0$	$0, 1$	$7, 0$	$-1, -1$
E	$2, 0$	$2, 0$	$2, 0$	$2, 0$	$3, 3$

Table 5: Impossibility under M3 rules with one-period memory

Suppose that an observation was in the minimal weak-curb set $\{A, B, C, D\} \times \{a, b, c, d\}$. Action E is a best response to some beliefs with the support $\{a, b, c, d\}$, namely near $(1/4)a * (1/4)b * (1/4)c * (1/4)d$, but it is never a best response to a belief over three or less actions in $\{a, b, c, d\}$. Since it is clearly not a best response to a pure action in $\{a, b, c, d\}$, consider a probability distribution over $\{a, b, c, d\}$ with three or two actions in the support. For example, a belief is of the form $pa * qb * (1 - p - q)c$. Then the best response is among $\{A, B, C\}$ since $\min_{0 \leq p+q \leq 1} \max\{7p, 7q, 7(1 - p - q)\} = 7/3 > 2$. Other cases are similar.

In this game, any M3 rule generates a belief with up to three pure actions in the support: the previous period action and two possible best responses by the opponent using an M2 rule. Therefore a learning process using only M3 rules does not leave the weak-curb set $\{A, B, C, D\} \times \{a, b, c, d\}$.

It is straightforward to extend this example for general finite level iteration and finite length memory case. See Proposition 2 and Section 4.

The above examples suggest the following.

(1) Minimal weak-curb sets can be reached under rather simple behavior rules, if there is a sufficient diversity of rules across two players. (2) Given a limited set of behavior rules, there are stage games with unique and strict Nash equilibrium, which no learning process can reach from some initial states. (3) The obstacle to get out of a non-curb set is the lack of feasible beliefs that rationalize an action outside of a weak-curb set under a limited set of behavior rules. The lack of rationalization of actions outside of a non-curb set is not only due to the limited ability to iterate best responses or mix them but also due to the payoff structure of the stage game.

3 Possibility and Impossibility of Learning with One-Period Memory

In this section we focus on very limited information case, i.e., one-period memory, and investigate the properties of long-term outcomes of general finite stage games, as the set of behavior rules vary.

3.1 Model

Let V_1, V_2 be populations of players. In each period $t = 1, 2, \ldots$, one player is drawn from each V_i to play the role of player i in a stage game. We call each player drawn from V_i player i. The stage game $G = (A_1, A_2, u_1, u_2)$ is a two-person normal-form game. The elements of A_i $(i = 1, 2)$ are called pure actions and we assume that A_i's are finite. The function $u_i : A_1 \times A_2 \to \Re$ is a payoff function for player $i = 1, 2$.

In each period $t = 1, 2, \ldots$, each player i chooses a pure action $a_i(t) \in A_i$ using a *behavior rule* and the information regarding the past action profiles. A behavior rule is a function from one's information to the set of actions A_i. We assume that players receive only the previous period action combination as the information. For each period $t = 1, 2, \ldots$, let $h_t = (a_1(t-1), a_2(t-1))$ be the previous period action combination. The initial observation $h_1 = (a_1(0), a_2(0))$ is arbitrary. The projection to i-th player's action is written as h_{it}.

To define behavior rules, we need some more set up. For a finite set X, let $\Delta(X)$ be the set of probability distributions over X. The set $\Delta(A_i)$ is the set of all mixed actions by player i. We extend the payoff function in the usual way to the expected payoff function $Eu_i : \Delta(A_1) \times \Delta(A_2) \to \Re$.

For each player $i = 1, 2$ and each mixed action $\sigma_j \in \Delta(A_j)$ by an opponent ($j \neq i$), define the set of pure-action best responses to σ_j as

$$BR_i(\sigma_j) = \{a_i \in A_i \mid Eu_i(a_i, \sigma_j) \geq Eu_i(x, \sigma_j) \ \ \forall x \in A_i\}.$$

For simplicity, assume that for each **pure** action $a_j \in A_j$ by the opponent, the pure best response $BR_i(a_j)$ is a singleton.[8] We call this assumption the *genericity assumption* of the stage game. By a slight abuse of notation, we also define the set of pure-action best responses to some mixed action in a set $X \subset \Delta(A_j)$;

$$BR_i(X) = \cup_{\sigma_j \in X} BR_i(\sigma_j).$$

In the following, we define the behavior rules. We describe the principle (or the reasoning) of a behavior rule first and then the functional form given a stage game.

First, we define a backward-looking behavior rule to choose according to the history.

S0 rule (conservative behavior rule):
Play the same action as in the previous period. The functional form is $a_i(t) = h_{it}$.

With one-period memory, there is a unique backward-looking behavior rule. For a generalization to multiple-period memory, see Section 4.

Second, we define *simple behavior rules* based on iterative best responses.

[8]The set of such games is of measure one in the space of finite two-person games.

S1 rule (adaptive behavior rule):

Play a best response to an opponent using S0 rule. The functional form is $a_i(t) = BR_i(h_{jt})$.

S2 rule (one-step forward-looking rule):

Play a best response to an opponent using S1 rule; $a_i(t) = BR_i(BR_j(h_{it}))$. This rule uses a forward-looking reasoning that the opponent reacts to your population's past action using the adaptive behavior rule. To use S2 rule, one needs to know the opponents' payoff function. Stahl [15] provides evidence that human subjects can compute a-few-times iterated best responses. Selten [12] considers a rule similar to S2 called anticipatory strategies.

S3 rule (two-step forward-looking rule):

Play a best response to an opponent using S2 rule; $a_i(t) = BR_i(BR_j(BR_i(h_{jt}))$.

And so on. The index number of the rules indicates the iteration of best responses.

One can define higher level S rules (many-step forward-looking behavior rules) which chooses a best response to the opponent using the one-step lower behavior rule. With one period memory and the genericity assumption, all simple behavior rules are single-valued.

A justification of our focus on the iterative best response is that the level of reasoning can be interpreted as the level of "sophistication". Another justification is experimental. The sequence of works by Stahl ([15], [16], [17] etc.) gives evidence that human subjects use similar iterative rules.

Third, players may want to allow multiple behavior rules by the opponent as possible, form a belief over the possible behavior rules, and play a best response to it.

M2 rules (mixed rules incorporating S0 and S1 rules by the opponent):

Play a best response to some probability distribution over S0 and S1 rules by the opponent. The functional form is $a_i(t) \in BR_i(\sigma_j)$ for some $\sigma_j \in \Delta(\{h_{jt}\} \cup BR_j(h_{it}))$.

In words, if a player believes that the next opponent uses either S0 rule or S1 rule, then his reaction falls in this group of behavior rules. The iteration of best responses is up to twice. An M2 rule is specified by the weight $w_1 \in [0, 1]$ on the possibility that the opponent uses S1 rule and denoted as $M2(w_1)$. The degenerate $M2(0)$ rule is equivalent to S1 rule and $M2(1)$ rule is equivalent to S2 rule.

M3 rules (mixed rules incorporating S0 and M2 rules by the opponent):
Play a best response to some probability distribution over the use of S0 and M2 rules (of various weights) by the opponent, i.e., $a_i(t) \in BR_i(\sigma_j)$ for some $\sigma_j \in \Delta(\{h_{jt}\} \cup BR_j(\Delta(\{h_{it}\} \cup BR_i(h_{jt}))))$.

An M3 rule specifies the weight on the use of S0 rule by the opponent and the weight distribution on the w_1 parameters of all possible M2 rules by the opponent.

Higher level mixed rules are similarly defined and include all lower level M rules and S rules as special-weight cases. Moreover, an Mn rule may not be single-valued, due to a mixed-action belief.

A *learning process* is an infinite (possibly stochastic) process of action profiles $\{(a_1(t), a_2(t))\}_{t=1}^{\infty}$, which is determined by an initial information $(a_1(0), a_2(0))$, a set of behavior rules B (with the information structure embedded), and a mechanism of how players choose a behavior rule from B in each period.

Finally we define possible limit action profiles of a learning process. A product set $C_1 \times C_2 \subset A_1 \times A_2$ is *closed under rational behavior* (a curb set) if $BR_1(\Delta(C_2)) \times BR_2(\Delta(C_1)) \subset C_1 \times C_2$. (Basu and Weibull [1].) That is, for any belief over the opponent's actions in C_j, the pure best response is contained in C_i.

A product set $C_1 \times C_2 \subset A_1 \times A_2$ is called a *weak-curb set* if it is closed under best response to itself; $BR_1(C_2) \times BR_2(C_1) \subset C_1 \times C_2$. A curb set is a weak-curb set but not vise versa (Example 4). Other related concepts are defined when they appear.

A product set $C_1 \times C_2 \subset A_1 \times A_2$ is *minimal in a property* X (called minimal-X set) if (i) it satisfies the property X, and (ii) there is no proper subset of $C_1 \times C_2$ which satisfies the property X. Since the entire action set $A_1 \times A_2$ is a curb (resp. weak-curb) set, a minimal-curb set (resp. minimal weak-curb set) exists for any finite game.

3.2 Convergence to a minimal weak-curb set

Def. A learning process $\{(a_1(t), a_2(t))\}_{t=1}^{\infty}$ *converges almost surely* to a set $X \subset A_1 \times A_2$ of action profiles if $P[\exists t^* < \infty; \quad (a_1(t), a_2(t)) \in X \quad \forall t \geq t^*] = 1$.

Proposition 1 *Assume that all players use only simple behavior rules or the conservative behavior rule and that there exists a lower bound $\epsilon > 0$, which is time and state independent, such that in each period, the probability is at least ϵ that one of the players uses an odd level rule (S1,S3,...) and the other uses an even level rule including S0 rule. Then, for any finite stage game $G = (A_1, A_2, u_1, u_2)$ and for any initial observation $(a_1(0), a_2(0)) \in A_1 \times A_2$, the learning process converges almost surely to one of the minimal weak-curb sets.*

Proof: See Appendix. The idea of the proof is described in Example 2.

The sufficient condition requires that different levels (odd and even) of behavior rules should be used at least with probability ϵ. It can be satisfied for example if there is ϵ probability that one player uses the adaptive rule and the other uses the conservative rule (there are two cases and both have probability at least ϵ). Hence the proposition encompasses a wide range of configurations of available behavior rules: from very limited set of behavior rules of only S0 and S1 rules to large sets of rules including many-step forward-looking behavior rules. It also includes processes in which players conduct rule-learning by adjusting behavior rules over time.

A question is whether and when the minimal weak-curb sets are plausible limit action sets. In some games, they coincide with Nash equilibria and thus even simple behavior rules can find Nash equilibria. We show two such classes of games.

Young [18] defines weakly acyclic games in which the deterministic adaptive process converges to a sequence of strict Nash equilibria (a convention) when the memory is long enough as compared to the sample size. For completeness of the paper we repeat his definition with our notation. Define the *best-reply graph* of a finite game G as follows: each vertex is an action combination $s \in A_1 \times A_2$, and for every two vertices s and s', there is a directed edge $s \to s'$ if and only if $s \neq s'$ and there exists exactly one player i such that s'_i is a best response to a_j and $s'_j = a_j$. A game G is *acyclic* if its best reply graph contains no directed cycles. It is *weakly*

acyclic if, from any initial vertex s, there exists a directed path to some vertex s^* from which there is no exiting edge (a sink).

A game is *weakly acyclic* if and only if from every action combination there exists a finite sequence of best responses by one player at a time that ends in a strict, pure Nash equilibrium. (Young, [18] p. 64.) Hence in weakly acyclic games, minimal weak-curb sets and minimal curb sets coincide and are the strict pure Nash equilibria. To compare with Young's no-mistake process, our model requires less and unperturbed information of the history but adds the probability of non-adaptive behavior rules.

Next, consider the supermodular games in Milgrom and Roberts [9].[9] We describe their definitions with our notation. Assume that the action set A_i $(i = 1, 2)$ comes with a partial order $\geq_i$. The set of action combinations $S = A_1 \times A_2$ is endowed with the product order $\geq$ that is, $(s_1, s_2) \geq (s_1', s_2')$ if and only if $a_i \geq s_i'$ for each $i = 1, 2$. When the game $G = (\{1, 2\}, A_1, A_2, u_1, u_2)$ is finite, the essential requirement[10] for G to be a *supermodular game* is the strategic complementarity: for each $i = 1, 2$, u_i has increasing differences in a_i and a_j; i.e., for all $a_i \geq a_i'$, the difference $u_i(a_i, a_j) - u_i(s_i', a_j)$ is nondecreasing in a_j.

In words, when the second player increases his choice variable(s), it becomes more profitable for the first to increase his as well. (Milgrom and Roberts, [9] p. 1261.)

Remark 1 *Let G be a supermodular game. Then minimal weak-curb sets, minimal curb sets, and pure Nash equilibria coincide.*

Proof: Suppose that $C_1 \times C_2$ is a minimal weak-curb set of G. Consider a restricted game $G' = (C_1, C_2, u_1, u_2)$. Then G' is also a supermodular game.

[9] Two-person supermodular games are related to potential games. See Brânzei et al. [2].

[10] To be precise, other conditions are
(A1) A_i is a complete lattice; i.e., for each two element set $\{x, y\} \subset A_i$, there is a supremum for $\{x, y\}$ (denoted $x \wedge y$) and an infimum (denoted $x \vee y$), and for all nonempty subsets $T \subset A_i$, $inf(T) \in A_i$ and $sup(T) \in S$.
(A2-1) u_i is order continuous in a_j (for fixed a_i); i.e., for each chain C (a totally ordered subset of A_j), it converges along C in both the increasing and decreasing directions, that is, $\lim_{a_j \in C, a_j \downarrow inf(C)} u_i(a_i, a_j) = u_i(a_i, inf(C))$ and $\lim_{a_j \in C, a_j \uparrow sup(C)} u_i(a_i, a_j) = u_i(a_i, sup(C))$.
(A2-2) $u_i : S \to \Re \cup \{-\infty\}$ is order upper semi-continuous in a_i (for fixed a_j); i.e.,
$\limsup_{a_i \in C, a_i \downarrow inf(C)} u_i(a_i, a_j) \leq u_i(inf(C), a_j)$ and $\liminf_{a_i \in C, a_i \uparrow sup(C)} u_i(a_i, a_j) \leq u_i(sup(C), a_j)$.
(A2-3) u_i has a finite upper bound;
(A3) u_i is supermodular in x_i (for fixed x_j); i.e., for all $x, y \in S$, $u_i(x) + u_i(y) \leq u_i(x \wedge y) + u_i(x \vee y)$.
It is easy to check that these are satisfied if G is finite.

Step 1: G' has a pure Nash equilibrium (s_1^*, s_2^*).

By Theorem 5 of Milgrom and Roberts [9]. //

Step 2: (s_1^*, s_2^*) is a Nash equilibrium of G.

Since $C_1 \times C_2$ is a weak-curb set, $BR_1(s_2^*) \subset C_1$. Since $u_1(s_1^*, s_2^*) \geq u_1(s_1, s_2^*)$ for all $s_1 \in C_1$, $s_1^* \in BR_1(s_2^*)$. Similarly, $s_2^* \in BR_2(s_1^*)$. //

By the minimality of $C_1 \times C_2$, we have that $C_1 \times C_2 = (s_1^*, s_2^*)$. ■

Supermodular games are generalized coordination games. Proposition 1 implies that players who are restricted to use only simple behavior rules can eventually coordinate actions under some diversity in thinking.

3.3 Nonconvergence to a Nash equilibrium

In games which possess minimal weak-curb sets that are not curb, processes with only simple behavior rules may get stuck in a non-curb set (Example 4.) Such games have cyclic best responses that are not easily escaped by simple reasoning. When the players can use mixed behavior rules and expand their beliefs, the process may move away from a non-curb set. Therefore it is natural to investigate whether learning processes using mixed behavior rules can discover minimal curb sets (instead of weak-curb sets). Hurkens [5] shows that mixed behavior rules up to infinite iteration of best responses are sufficient for convergence to a minimal curb set.

By contrast, if players are restricted to use Mn rules for some finite n, we found a class of finite-action stage games, in which any learning process using up to Mn rules cannot reach a strict Nash equilibrium (a singleton minimal curb set). The intuition is described in Example 5 in Section 2.

Proposition 2 *If players use only Mn rules for some finite n and the conservative behavior rule, then there exist a class of finite stage games with a unique Nash equilibrium, in which no learning process can reach the Nash equilibrium from some initial observation.*

Proof: Consider a class of games of the following form, where $x > 0$, $z > 0$ and $x/n > y > x/(n+1)$.

P1 \ P2	1	2	3	$\cdots$	n	n+1	n+2
1	$x, 0$	$0, x$	$0, 0$	$\cdots$	$0, 0$	$0, 0$	$-1, -1$
2	$0, 0$	$x, 0$	$0, x$	$\cdots$	$0, 0$	$0, 0$	$-1, -1$
3	$0, 0$	$0, 0$	$x, 0$	$\cdots$	$0, 0$	$0, 0$	$-1, -1$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	$0, 0$	$0, 0$	$0, 0$	$\cdots$	$x, 0$	$0, x$	$-1, -1$
n+1	$0, x$	$0, 0$	$0, 0$	$\cdots$	$0, 0$	$x, 0$	$-1, -1$
n+2	$y, 0$	$y, 0$	$y, 0$	$\cdots$	$y, 0$	$y, 0$	z, z

Table 6: A cyclic game with a unique Nash equilibrium

The inequality $x/n > y$ implies that $\{1, \ldots, n+1\} \times \{1, \ldots, n+1\}$ does not allow a belief with n actions or less in the support which has a best response outside of it. The inequality $y > x/(n+1)$ implies that there are beliefs with $n+1$ actions in the support which has a best response outside of $\{1, \ldots, n+1\} \times \{1, \ldots, n+1\}$. Hence if a process using only Mn rules and the conservative behavior rule enters the non-curb set $\{1, \ldots, n+1\} \times \{1, \ldots, n+1\}$, it cannot leave and reach the unique Nash equilibrium $\{n+2\} \times \{n+2\}$. ■

The essence of the proof is that it is possible to construct a stage game with a very large cycle of best responses so that players with limited memory and reasoning cannot form a belief that rationalizes an action outside of the cycle.

3.4 Limit actions under mixed behavior rules

The proof of the impossibility result (Proposition 2) uses a game with a minimal weak-curb set which cannot be left by beliefs with a limited size support. A conjecture arises that it is the size of the support of feasible beliefs that determines the limit set of actions. If so, the limit actions are solely dependent on the feasible behavior rules. We show that in a subclass of stage games this conjecture is true and a counter example for other games.

A product set $C_1 \times C_2$ is *closed under beliefs with the support of two or less actions (a C(2)-set)* if for any state $(a_1, a_2) \in C_1 \times C_2$, $BR_i(\Delta_2(C_j)) \subset C_i$ for each $i = 1, 2$, where $\Delta_2(C_j) = \{\sigma \in \Delta(C_j) \mid |supp(\sigma)| \leq 2\}$.

A curb set is a $C(2)$-set and a $C(2)$-set is a weak-curb set.

Proposition 3 *Take an arbitrary finite stage game G such that for each $a_i \in A_i$, there exists $a_j \in A_j$ $(j \neq i)$ such that $a_i = BR_i(a_j)$, for $i = 1, 2$. Assume that all players use only M2 rules and the conservative behavior rule and that there exists a lower bound $\epsilon > 0$, which is time and state independent, such that in each period, the probability is at least ϵ that each of S0, S1 and S2 rules are used and the probability density function of M2 rules with weight $w_1 \in (0, 1)$ (which can be time and state dependent) is bounded away from zero by*[11] *ϵ. Then, for any initial observation $(a_1(0), a_2(0)) \in A_1 \times A_2$, the learning process converges almost surely to a minimal $C(2)$-set.*

Proof: See Appendix.

Proposition 3 implies that for processes using only M2 rules and S0 rule, the limit actions are characterized solely by the nature of the feasible behavior rules, i.e., the size of the support of feasible beliefs. This is a remarkable implication.

The assumption on the stage game that each pure action is a best response to some pure action by the opponent makes the best response function a bijection. It is necessary for a process to move out of a non-C(2)-set. Consider the game in Table 7 and any learning process using only M2 rules and S0 rule.

P1 \ P2	a	b	c	d	e	f
A	$7, 0$	$1, 0$	$0, 4$	$0, 7$	$0, 0$	$-1, -1$
B	$0, 4$	$0, 0$	$0, 4$	$1, 0$	$0, 7$	$-1, -1$
C	$0, 4$	$0, 7$	$7, 0$	$0, 0$	$1, 0$	$-1, -1$
D	$4, 0$	$0, 0$	$4, 0$	$0, 0$	$0, 0$	$10, 10$

Table 7

Consider the product set $\{A, B, C\} \times \{a, b, c, d, e\}$. This set is not closed under some beliefs over $\{a, c\}$ but no M2 rule by Pop 1 player can generate such belief, since none of a or c is the action of the adaptive behavior rule. Hence although $\{A, B, C\} \times \{a, b, c, d, e\}$ is not closed under belief with two actions in the support, it is closed under M2 rules (and S0 rule).

[11] A function $f : X \to \Re_+$ is bounded away from zero by ϵ if $f(x) \geq \epsilon$ for any $x \in X$.

An interesting implication of this example is that if players use only simple behavior rules, the process could stay in a minimal weak-curb set $\{A, B, C\} \times \{b, d, e\}$, while adding mixed rules enlarged the limit set. Hence it is not always useful to introduce more behavior rules if the purpose was to nail down the set of long-term outcomes. In other words, smartness alone cannot find a rational action without the aid of a payoff incentive to explore actions.

Proposition 3 does not extend to the processes using Mn rules with $n \geq 3$. Consider the game in Table 8 and any learning process using only M3 rules and S0 rule. This game satisfies the bijection property of the best response function.

P1 \	a	b	c	d	e	f	g
A	$7,0$	$0,1$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$
B	$0,0$	$7,0$	$0,1$	$0,0$	$0,0$	$0,0$	$0,0$
C	$0,0$	$0,0$	$7,0$	$0,1$	$0,0$	$0,0$	$0,0$
D	$0,0$	$0,0$	$0,0$	$7,0$	$0,1$	$0,0$	$0,0$
E	$0,0$	$0,0$	$0,0$	$0,0$	$7,0$	$0,1$	$0,0$
F	$0,1$	$0,0$	$0,0$	$0,0$	$0,0$	$7,0$	$0,0$
G	$3,0$	$0,0$	$3,0$	$0,0$	$3,0$	$0,0$	$2,2$

Table 8

One of the minimal weak-curb sets, $\{A, B, \ldots, F\} \times \{a, b, \ldots, f\}$, is not closed under some beliefs near $(1/3)a * (1/3)c * (1/3)e$ for Pop 1 player. Although a Pop 1 player using an M3 rule can have a belief with three actions in the support, an "exit state" $(a_1^*, a_2^*) \in \{A, B, \ldots, F\} \times \{a, b, \ldots, f\}$ from which a Pop 1 player can play action G must satisfy the following property: Let $a_2^{(1)} = a$, $a_2^{(2)} = c$ and $a_2^{(3)} = e$. In this game, (a_1^*, a_2^*) is an "exit state" if and only if there exists a permutation $\pi : \{1, 2, 3\} \to \{1, 2, 3\}$ such that

$$\begin{aligned} a_2^{\pi(1)} &= a_2^* \\ a_2^{\pi(2)} &= BR_2(a_1^*) \\ a_2^{\pi(3)} &= BR_2(BR_1(a_2^*)). \end{aligned}$$

Namely, these iterated best responses must coincide with $\{a, c, e\}$. For the game of Table 9 this is impossible because, for any $a_2 \in \{a, c, e\}$, if we let $a_2^* = a_2$, then $BR_2(BR_1(a_2^*)) \notin \{a, c, e\}$.

Therefore, in general, the limit actions of learning processes with a limited set of behavior rules are not characterized by the number of the actions that one can include in his beliefs (which is solely determined by the behavior rules). We have to consider the structure of the best responses which depends on the stage game.

4 Longer Memory Model

The one-period memory model is not so restrictive as it may appear since the definition of a time period only restricts that during a period no player can alter actions. For example, teachers can give grades of a course only once in a semester and a government can determine the budget only once a year. Hence the one-period model can encompass a wide range of interactive problems of various length in real time.

It is, however, of some interest to extend the model for multiple-period memory and see whether and how the above results change. An easy observation is that one can always ignore more information than the most recent period and can use the same set of behavior rules of the one-period memory case. Therefore the convergence to minimal weak-curb sets (Proposition 1) in the one-period model holds with degenerate behavior rules of multi-period memory.

We describe how impossibility result of Proposition 2 can be extended for multi-period memory model. The longer memory contributes to larger support of feasible beliefs, but as long as the memory is finite and the iterative reasoning is finite, the feasible beliefs have only a finite support. Hence the same type of stage games in Proposition 2 prevents the process to move out of a non-curb set.

With multiple-period memory, a conservative behavior rule chooses one of the observed past actions in one's own population. Since there can be multiple observed actions, there is room for another criterion how to choose among them. There are many conservative behavior rules including the popular "performance-based" rule, which chooses the highest payoff action in the past. The range of possible actions by some conservative behavior rule is, however, exactly the span of the observed actions. Therefore, if m periods in the past can be observed, the maximal number of different actions that can be chosen by some conservative behavior rule is m. (This holds under random sampling model as well by defining that m is the size of sample instead of the memory.)

Consider Young's adaptive rule, which assigns a best response to one of the actions in the observed history (or in the sample from the history).[12] An adaptive behavior rule can also have up to m different actions under the genericity assumption. An M2 rule specifies a probability distribution over at most $2m$ possible actions by conservative and adaptive rules and assigns a best response to it. While the number of pure-action best responses to mixed actions may exceed $2m$, one can construct a stage game similar to Table 6 to restrict them to $2m$ as well. Any higher level mixed rule can be also restricted to have beliefs with finite support. Therefore it is possible to construct a game such that a weak-curb set cannot be left with some finite number of actions in the support of a belief, in the spirit of Proposition 2. Hence Proposition 2 can be extended to multiple-period memory.

For the intuition, consider the two-period memory case with the following stage game.

P1 \ P2	a	b	c	d	e	f
A	$x, 0$	$0, 1$	$0, 0$	$0, 0$	$0, 0$	$-1, -1$
B	$0, 0$	$x, 0$	$0, 1$	$0, 0$	$0, 0$	$-1, -1$
C	$0, 0$	$0, 0$	$x, 0$	$0, 1$	$0, 0$	$-1, -1$
D	$0, 0$	$0, 0$	$0, 0$	$x, 0$	$0, 1$	$-1, -1$
E	$0, 1$	$0, 0$	$0, 0$	$0, 0$	$x, 0$	$-1, -1$
F	$y, 0$	$y, 0$	$y, 0$	$y, 0$	$y, 0$	$1, 1$

Table 9

Let $x/4 > y > x/5 > 0$.

The conservative behavior rule with two period memory prescribes to play one of the two observed actions of one's own population. Hence $a_i(t) \in \{a_i(t-2), a_i(t-1)\}$ for a Pop i player. The adaptive behavior rule prescribes a best response to one of the observed actions by the opponent population. Hence $a_i(t) \in \{BR_i(a_j(t-2)), BR_i(a_j(t-1))\}$ for a Pop i player, where $j \neq i$. A M2 rule of Pop i prescribes a best response to a probability distribution over $\{a_j(t-2), a_j(t-1), BR_j(a_i(t-2)), BR_j(a_i(t-1))\}$. Therefore beliefs under M2 rules have at most four pure actions in the support. If the observation of the previous two periods is

[12] With multiple-period memory, there are other definitions of adaptiveness such as best response to the empirical distribution of actions.

confined in the non-curb set $\{A, B, C, D, E\} \times \{a, b, c, d, e\}$, then the resulting action profile does not leave it. For example if (A, c) and (C, e) are observed, a feasible belief by Pop 1 player is a probability distribution over $\{c, e, BR_2(A), BR_2(C)\} = \{b, c, d, e\}$ and the range of best responses is $\{B, C, D, E\}$.

References

[1] K. Basu, J. Weibull, Strategy Subsets Closed under Rational Behavior, Economics Letters, 36 (1991), 141-146.

[2] R. Brânzei, L. Mallozi, S. Tijs, Supermodular Games and Potential Games, Journal of Mathematical Economics, 39 (2003), 39-49.

[3] D. Fudenberg, D. Levine, Self-Confirming Equilibrium, Econometrica, 61 (1993), 523-545.

[4] D. Fudenberg, D. Levine, The Theory of Learning in Games, MIT press, Boston, 1998.

[5] S. Hurkens, Learning by Forgetful Players, Games and Economic Behavior, 11 (1995), 304-329.

[6] J. Josephson, Stochastic Adaptation in Finite Games Played by Heterogeneous Populations, manuscript, Stockholm School of Economics, 2002.

[7] R. Marimon, E. McGrattan, On Adaptive Learning in Strategic Games, in Learning and Rationality in Economics. Kirman and Salmon eds. Basil Blackwell, Cambridge, 1995.

[8] A. Matros, Clever Agents in Adaptive Learning, Journal of Economic Theory, 111 (2003), 110-124.

[9] P. Milgrom, J. Roberts, Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities, Econometrica, 58 (1990), 1255-1277.

[10] P. Milgrom, J. Roberts, Adaptive and Sophisticated Learning in Normal Form Games, Games and Economic Behavior, 3 (1991), 82-100.

[11] A. Roth, V. Prasnikar, M. Okuno-Fujiwara, S. Zamir, Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study, American Economic Review, 81 (1991), 1068-1095.

[12] R. Selten, Anticipatory Learning in 2 Person Games. In: Selten R (ed.) Game Equilibrium Models I, Springer Verlag, 1991.

[13] R. Selten, Features of Experimentally Observed Bounded Rationality, European Economic Review, 42 (1998), 413-436.

[14] D. Stahl, Evolution of Smart$_n$ Players, Games and Economic Behavior 5 (1993), 604-617.

[15] D. Stahl, Boundedly Rational Rule Learning in a Guessing Game, Games and Economic Behavior 16 (1996), 303-330.

[16] D. Stahl, Rule Learning in Symmetric Normal-Form Games: Theory and Evidence, Games and Economic Behavior 32 (2000), 105-138.

[17] D. Stahl, P. Wilson, Experimental Evidence on Players' Models of Other Players, Journal of Economic Behavior and Organization, 25 (1994), 309-327.

[18] P. Young, The Evolution of Conventions, Econometrica, 61 (1993),57-84.

APPENDIX

Proof of Proposition 1: Without loss of generality, assume that $A_1 \times A_2$ is not a minimal weak-curb set. Since $A_1 \times A_2$ is a weak-curb set, it suffices to prove that from any initial observation $(a_1, a_2) \in A_1 \times A_2$, there is a positive probability that the action combination enters a smaller weak-curb set in a finite number of periods. Then the same logic can be applied to each of the smaller weak-curb sets, which is not minimal, until the action combination reaches a minimal weak-curb set. Since $A_1 \times A_2$ is finite, we reach a minimal weak-curb set in a finite number of periods with a positive probability.

Let $W_1^1 \times W_2^1, W_1^2 \times W_2^2, \ldots, W_1^K \times W_2^K$ be the largest[13] weak-curb sets, which are proper subsets of $A_1 \times A_2$. We call them "smaller weak-curb sets" (than $A_1 \times A_2$).

Step 1: A state $(a_1, a_2) \in A_1 \times A_2$ belongs to one of the following sets:
(a) $\cup_{k=1}^K [W_1^k \times W_2^k]$ which is the set of states in some smaller weak-curb set,
(b) $\left[\cup_{k=1}^K [(A_1 \setminus W_1^k) \times W_2^k]\right] \cup \left[\cup_{k=1}^K [W_1^k \times (A_2 \setminus W_2^k)]\right]$ which is the set of states such that one action belongs to a smaller weak-curb set, and
(c) $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ which is the set of states such that no action belongs to any of the smaller weak-curb set.

Proof:

Case 1: Suppose that $a_1 \in W_1^k$ for some $k \in \{1, 2, \ldots, K\}$.

If $a_2 \in W_2^k$, then $(a_1, a_2) \in \cup_{k=1}^K [W_1^k \times W_2^k]$. If $a_2 \in A_2 \setminus W_2^k$, then $(a_1, a_2) \in \cup_{k=1}^K [W_1^k \times (A_2 \setminus W_2^k)]$.

Case 2: Suppose that $a_1 \notin W_1^k$ for any $k \in \{1, 2, \ldots, K\}$, i.e., $a_1 \in A_1 \setminus \cup_k W_1^k$.

If $a_2 \in W_2^j$ for some $j \in \{1, 2, \ldots, K\}$, then $(a_1, a_2) \in \cup_{k=1}^K [(A_1 \setminus W_1^k) \times W_2^k]$. If $a_2 \notin W_2^k$ for any $k \in \{1, 2, \ldots, K\}$, then $(a_1, a_2) \in (A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$. //

Note that the set (b) may have states such that one action belongs to a smaller weak-curb set W_i^k while the other action belongs to another weak-curb set W_j^m. See Figure 1.

=====Insert Figure 1 about here. =====

[13] weak-curb sets can be nested but cannot overlap.

Note also that $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ is not a weak-curb set since the W_i^k's cover all smaller weak-curb sets of $A_1 \times A_2$.

Step 2: If a process enters a smaller weak-curb set $W_1^k \times W_2^k$ for some $k \in \{1, 2, \ldots, K\}$, then it will not get out.

Proof: Take any state $(a_1, a_2) \in W_1^k \times W_2^k$ for some $k \in \{1, 2, \ldots, K\}$. If Pop i player uses S0 rule, the next period action is $a_i(t) = a_i \in W_i^k$. If Pop i player uses S1 rule, the next period action is $BR_i(a_j) \in W_i^k$ by the definition of the weak-curb set. If Pop i player uses S2 rule, the next period action is $BR_i(BR_j(a_i)) \in W_i^k$... Any iterative best response stays in the same set. Therefore the process does not get out of $W_1^k \times W_2^k$. //

Step 3: If a process enters $\left[\cup_{k=1}^K [(A_1 \setminus W_1^k) \times W_2^k]\right] \cup \left[\cup_{k=1}^K [W_1^k \times (A_2 \setminus W_2^k)]\right]$, i.e., one of the observed action belongs to a smaller weak-curb set W_i^k for some $i \in \{1, 2\}$ and some $k \in \{1, 2, \ldots, K\}$, then the process enters $W_1^k \times W_2^k$ in the next period with a probability not less than ϵ.

Proof: Let (a_i, a_j) be the observation and $a_i \in W_i^k$ for some $k \in \{1, 2, \ldots, K\}$.

By the assumption, the probability that Pop i player uses an **even** level simple behavior rule and Pop j player uses an **odd** level rule at the same time is not less than ϵ.

In this case the next period action of Pop i player is either a_i itself (using S0 rule) or an iterative best response of $BR_j(a_i) \in W_j^k$. That is, if he uses S2 rule, $a_i(t) = BR_i \circ BR_j(a_i)$. If he uses S4 rule, $a_i(t) = BR_i \circ BR_j \circ BR_i \circ BR_j(a_i)$, and so on. By the definition of weak-curb set, all these iterative best responses belong to the same W_i^k. Similarly, the next period action of Pop j player is an iterative best response of a_i, i.e., $a_j(t) \in \{BR_j(a_i), BR_j \circ (BR_i \circ BR_j)(a_i), \ldots, BR_j \circ (BR_i \circ BR_j)^n(a_i), \ldots\} \subset W_j^k$. //

Step 4: Suppose that the observation (a_1, a_2) belongs to $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$. There exists a probability $p > 0$, independent of time and state, and a finite period $\bar{t}$ such that the process gets out of this set within $\bar{t}$ periods with probability at least p.

Proof: Recall that the product set $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ is not a weak-curb set. Hence there exists $i \in \{1, 2\}$ and an "exit observation" $\bar{a}_i \in (A_i \setminus \cup_k W_i^k)$ such that $BR_j(\bar{a}_i) \notin (A_j \setminus \cup_k W_j^k)$ or equivalently $BR_j(\bar{a}_i) \in W_j^k$, for some $k \in \{1, 2, \ldots, K\}$. This means that if $\bar{a}_i$ is observed, the

best response by Pop j is outside of the set $(A_j \setminus \cup_k W_j^k)$. Collect these "exit observations" for both populations since there can be many. Let $T_i = (A_i \setminus \cup_k W_i^k)$ and for any $A_1 \times A_2 \subset A_1 \times A_2$, define $EXIT(A_i) := \{\bar{a}_i \in A_i \mid BR_j(\bar{a}_i) \notin A_j\}$ for $i = 1, 2$ and $j \neq i$.

Step 4-1: If one of the observed action $a_i \in EXIT(T_i)$.

Then Pop j player can use an **odd** level simple behavior rule to play an iterative best response to a_i of the form $BR_j \circ (BR_i \circ BR_j)^{n-1}(a_i)$ for some $n = 1, 2, \ldots$. All these actions belong to some weak-curb set W_j^k and hence the process gets out of $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ in one period.

Step 4-2: If none of the observations belongs to $EXIT(T_i)$ for some $i = 1, 2$. Then the observation (a_1, a_2) belongs to a smaller set $[T_1 \setminus EXIT(T_1)] \times [T_2 \setminus EXIT(T_2)]$. Let $T_i^2 = [T_i \setminus EXIT(T_i)]$ for $i = 1, 2$. Note that $T_1^2 \times T_2^2$ is not a weak-curb set. Therefore at least one of $EXIT(T_i^2)$ $i = 1, 2$ has an element. If one of the observation action a_i belongs to $EXIT(T_i^2)$, then by the same logic in Step 4-1, the process gets out of $T_1^2 \times T_2^2$.

We can continue the same argument but since the stage game is finite, after some $\bar{t} < \infty$ steps, all the states in $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ are exhausted. Then a process can get out of $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ in $\bar{t}$ periods. In each step, the probability of going to one of the previous steps is at least ϵ and thus the probability of getting out entirely from $(A_1 \setminus \cup_k W_1^k) \times (A_2 \setminus \cup_k W_2^k)$ is at least $p = \epsilon^{\bar{t}}$. //

This completes the proof of Proposition 1. ■

Proof of Proposition 3:

Step 1: Each weak-curb set $W_1 \times W_2$ is a "square set", i.e., $|W_1| = |W_2|$.

Proof: Let $W_1 \times W_2$ be a weak-curb set. By the genericity assumption, for each $a_2 \in W_2$, there is a unique $a_1 \in A_1$ such that $a_1 = BR_1(a_2)$. Therefore $|\{a_1 \in A_1 \mid a_1 = BR_1(a_2) \quad \exists a_2 \in W_2\}| = |BR_1(W_2)| = |W_2|$. Analogously, $|BR_2(W_1)| = |W_1|$. Suppose that $|W_1| > |W_2|$, then

$$|BR_2(W_1)| = |W_1| > |W_2| = |BR_1(W_2)|.$$

However, $BR_2(W_1) \subset W_2$ implies that $|BR_2(W_1)| \leq |W_2|$, a contradiction. //

Step 2: For any $i = 1, 2$ and any $a_i \in A_i$, there is a unique $a_j \in A_j$ such that $a_i = BR_i(a_j)$. (Hence BR_i is a bijection.)

Proof: Under the assumption of the stage game, for any $a_i \in A_i$, there exists at least one $a_j \in A_j$ such that $a_i = BR_i(a_j)$. Suppose that for some $a_i \in A_i$, there exist multiple $a'_j \neq a_j$ such that $a_i = BR_i(a_j)$ and $a_i = BR_i(a'_j)$. Then $|A_j| > |A_i|$, a contradiction to Step 1 (since $A_1 \times A_2$ is a weak-curb set). //

Step 3: Let $W_1 \times W_2$ be a minimal weak-curb set. Take an arbitrary $a_2 \in W_2$ and define sequences of best responses as follows:

$$\begin{aligned} a_1^{(1)} &:= BR_1(a_2), \\ a_2^{(1)} &:= BR_2(a_1^{(1)}), \\ &\cdots \\ a_1^{(t)} &:= BR_1(a_2^{(t-1)}), \\ a_2^{(t)} &:= BR_2(a_1^{(t)}), \\ &\cdots \end{aligned}$$

and let T be the smallest number such that $a_1^{(T+1)} \in \{a_1^{(1)}, \ldots, a_1^{(T)}\}$. Then $\{a_2^{(1)}, \ldots, a_2^{(T)}\} = W_2$

Proof: Since BR_i is a bijection, T is also the smallest number such that $a_2^{(T+1)} \in \{a_2^{(1)}, \ldots, a_2^{(T)}\}$. Let $W'_2 := \{a_2^{(1)}, \ldots, a_2^{(T)}\}$ and $W'_1 := \{a_1^{(1)}, \ldots, a_1^{(T)}\}$, then $W'_2 \subset W_2$ and $W'_1 \subset W_1$. Suppose that W'_2 is a proper subset of W_2. Since BR_i is a bijection, W'_1 is also a proper subset of W_1 and $W'_1 \times W'_2$ is a weak-curb set, a contradiction. //

Let $\{W_1^1 \times W_2^1, W_1^2 \times W_2^2, \ldots, W_1^k \times W_2^k\}$ be the collection of all <u>minimal</u> weak-curb sets of the stage game.

Step 4: Minimal weak-curb sets do not have an intersection.

Proof: Suppose that $[W_1^k \times W_2^k] \cap [W_1^m \times W_2^m] \neq \emptyset$ for some k, m. Then for any $(a_1, a_2) \in [W_1^k \times W_2^k] \cap [W_1^m \times W_2^m]$, $BR_i(a_j) \in W_i^k \cap W_i^m$ for each $i = 1, 2$ and $j \neq i$. Hence $BR_1(W_2^k \cap W_2^m) \times BR_2(W_1^k \cap W_1^m) \subset [W_1^k \cap W_1^m] \times [W_2^k \cap W_2^m]$, i.e., $[W_1^k \cap W_1^m] \times [W_2^k \cap W_2^m]$ is a weak-curb set, which is a contradiction. //

Step 5: For each player $i = 1, 2$, the projections of the minimal weak-curb sets, $\{W_i^1, \ldots, W_i^k\}$, partition the entire action set A_i. That is, the set $\tilde{A} = [A_1 \setminus (\cup_{k=1}^K W_1^k)] \times [A_2 \setminus (\cup_{k=1}^K W_2^k)]$ is empty. (See Figure 2.)

Proof: Suppose not. Since $\tilde{A}$ cannot be a weak-curb set, there exists $i = 1, 2$ and $a_i \in [A_i \setminus (\cup_{k=1}^K W_i^k)]$ such that $BR_j(a_i) \notin [A_j \setminus (\cup_{k=1}^K W_j^k)]$, i.e., $BR_j(a_i) \in W_j^k$ for some minimal weak-curb set $W_1^k \times W_2^k$. Let $a_j := BR_j(a_i)$. Since $W_1^k \times W_2^k$ is a weak-curb set, there exists $a_i' \in W_i^k$ such that $a_j = BR_j(a_i')$, a contradiction to Step 2. //

=====Insert Figure 2 about here. =====

Step 6: For each minimal C(2)-set $C_1 \times C_2$, there exists a subset $N \subset \{1, 2, \ldots, K\}$ of the indices of minimal weak-curb sets such that $C_i = \cup_{k \in N} W_i^k$ for each $i = 1, 2$.

Proof: By the definition, a C(2)-set is a weak-curb set and thus C_i cannot be a proper subset of some (single) W_i^k. Suppose that there exists W_i^k such that there exist $a_i \in C_i \cap W_i^k$ and $a_i' \in W_i^k \setminus C_i$. By a similar logic to the one in Step 4, $[C_1 \cap W_1^k] \times [C_2 \cap W_2^k]$ is a weak-curb set, which is a contradiction that $W_1^k \times W_2^k$ is a minimal weak-curb set. //

Step 6 implies that the entire action set $A_1 \times A_2$ is partitioned into three categories: the action combinations that belong to a minimal C(2)-set, those with one of the actions in a minimal C(2)-set, and those with none of the actions in a C(2)-set. (See Figure 2.) By a similar logic to the one in Proposition 1, if one of the actions belongs to a minimal C(2)-set, then with a positive probability the next period action profile is in the C(2)-set. If the process enters one of the minimal C(2)-sets, it stays there since no M2 rule would prescribe an action outside of a C(2)-set. Hence it suffices to show that if the process enters the product set of states in which none of the actions belongs to a minimal C(2)-set, then the process gets out of it. We show this in three additional steps.

Step 7: Let $\overline{C}_1 \times \overline{C}_2$ be a non-C(2)-set such that $\overline{C}_1 \times \overline{C}_2 = [\cup_{k \in N'} W_1^k] \times [\cup_{k \in N'} W_2^k]$ for some $N' \subset \{1, 2, \ldots, K\}$. There is an "exit state" $(a_1, a_2) \in \overline{C}_1 \times \overline{C}_2$, $i \in \{1, 2\}$, and $w_1 \in [0, 1]$ such that $BR_i(w_1 BR_j(a_i) * (1 - w_1)a_j) \notin \overline{C}_i$.

Proof: Since $\overline{C}_1 \times \overline{C}_2$ is not a C(2)-set, there exist $j \in \{1,2\}$ and (not necessarily distinct) $a_j, a'_j \in \overline{C}_j$ such that $BR_i(w_1 a_j * (1-w_1)a'_j) \notin \overline{C}_i$ for some $w_1 \in [0,1]$. By the assumption of the stage game, there exists $a_i \in A_i$ such that $a_j = BR_j(a_i)$. Moreover, $\overline{C}_1 \times \overline{C}_2 = [\cup_{k\in N} W_1^k] \times [\cup_{k\in N} W_2^k]$ and $a_j \in \overline{C}_j$ imply that this $a_i \in \overline{C}_i$. //

Step 8: There exists $p > 0$, which is time and state independent, such that the probability is at least p that from any exit state (a_1, a_2) the process gets out of $\overline{C}_1 \times \overline{C}_2$ in one period.

Proof: We have assumed that the density function f over w_1 is bounded away from zero by $\epsilon > 0$. For any exit state (a_1, a_2), there exists a player i and an interval $[\underline{w}_1, \overline{w}_1]$ such that $BR_i(w_1 BR_j(a_i) * (1-w_1)a_j) \notin \overline{C}_i$ if and only if $w_1 \in [\underline{w}_1, \overline{w}_1]$. Then the probability is at least $\epsilon(\overline{w}_1 - \underline{w}_1)$ that the process get out of $\overline{C}_1 \times \overline{C}_2$ by some M2 rule with parameter in the interval $[\underline{w}_1, \overline{w}_1]$. Finally take the minimum of $\epsilon(\overline{w}_1 - \underline{w}_1)$ over all exit states (which are finite). //

Step 9: Exit states in a non-C(2)-set $\overline{C}_1 \times \overline{C}_2 = [\cup_{k\in N} W_1^k] \times [\cup_{k\in N} W_2^k]$ are reachable from any other state in $\overline{C}_1 \times \overline{C}_2$ in a finite number of periods with a positive probability.

Proof: Take an arbitrary exit state. If it is in a minimal weak-curb set, by the same logic in Proposition 1, it can be reached. Suppose that the exit state does not belong to any of the minimal weak-curb set within $\overline{C}_1 \times \overline{C}_2$. Then it belongs to the "off-diagonal" area $S := \{(a_1, a_2) \in W_1^k \times W_2^m \mid \exists\ k, m \in N;\ k \neq m\}$. (See Figure 3.) Since each minimal weak-curb set within $\overline{C}_1 \times \overline{C}_2$ is not a C(2)-set, from any state in a minimal weak-curb set within $\overline{C}_1 \times \overline{C}_2$, the process can get out to S with a positive probability as in Step 7. Hence it suffices to prove that states in S are mutually reachable.

The "off-diagonal" area S can be divided into two sets $S_+ := \{(a_1, a_2) \in W_1^k \times W_2^m \mid \exists\ k, m \in N;\ k > m\}$ and $S_- := \{(a_1, a_2) \in W_1^k \times W_2^m \mid \exists\ k, m \in N;\ k < m\}$. (The indices of minimal weak-curb sets can be ordered according to the action order. See Figure 3.)

=====Insert Figure 3 about here. =====

It suffices to prove (1) that states within S_+ or S_- are mutually reachable and (2) that there exists a positive probability that from any state in S_+, a state in S_- is reached in one period and vice versa.

(1) Without loss of generality we only prove that all states in S_+ are mutually reachable within a finite number of periods with a positive probability. It suffices to prove the following: Take an arbitrary state $(a_1, a_2) \in S_+$ as the starting state. For any $a_2' \in \overline{C}_2$ such that $(a_1, a_2') \in S_+$ and any $a_1' \in \overline{C}_1$ such that $(a_1', a_2) \in S_+$, the action combination (a_1, a_2') and (a_1', a_2) are played within a finite number of periods with a positive probability. These states cover all horizontal and vertical movements from (a_1, a_2) in S_+ with possible overlap. By the symmetry, we only prove that from an arbitrary $(a_1, a_2) \in S_+$, any $a_2' \in \overline{C}_2$ such that $(a_1, a_2') \in S_+$ is played.

Suppose that Pop 1 player uses S0 rule and Pop 2 player uses S2 rule (M2 rule with parameter $w_1 = 1$), each of which occurs with probability at least $\epsilon > 0$. Take any $a_2 \in \overline{C}_2$ such that $(a_1, a_2) \in S_+$. Then there exist k, m such that $(a_1, a_2) \in W_1^k \times W_2^m$. It follows that $BR_1(a_2) \in W_1^m$ and $a_2^{(1)} := BR_2(BR_1(a_2)) \in W_2^m$. (See Figure 4.) Hence $(a_1, a_2^{(1)}) \in S_+$ can be reached in one period with probability at least ϵ^2.

=====Insert Figure 4 about here. =====

By repeatedly using S0 and S2 rules, $(a_1, a_2^{(2)}) \in S_+$ can be reached in two periods with probability at least ϵ^4, where $a_2^{(2)} := (BR_2 \circ BR_1)^2(a_2) \in W_2^m$. We can iterate this argument. Let T be the smallest number such that $a_2^{(T+1)} \in \{a_2^{(1)}, \ldots, a_2^{(T)}\}$, where $a_2^{(t)} := (BR_2 \circ BR_1)^t(a_2) = BR_2 \circ BR_1(a_2^{(t-1)})$ for each t. Since the action sets are finite, there exists such $T < \infty$.

By Step 3, we have that $\{a_2^{(1)}, \ldots, a_2^{(T)}\} = W_2^m$ and hence all states in $\{a_1\} \times W_2^m$ are reached within T periods with probability at least $\epsilon^{2T} > 0$. We can repeat this argument for any $m' < k$ such that $\{a_1\} \times W_2^{m'} \subset S_+$. Therefore all states of the form (a_1, a_2') are played within a finite number of periods with a positive probability. //

(2) Take an arbitrary state $(a_1, a_2) \in S_+$. There exist $k > m$ such that $a_1 \in W_1^k$ and $a_2 \in W_2^m$. Using the property of weak-curb sets, we have that $BR_2(a_1) \in W_2^k$ and $BR_1(a_2) \in W_1^m$, hence if both players use the adaptive rule (which occurs with at least probability ϵ), the process enters S_- in one period. The opposite is similar. //

Steps 7-9 imply that the process gets out of non-C(2)-set within a finite number of periods with a positive probability. This completes the proof of Proposition 3. ■

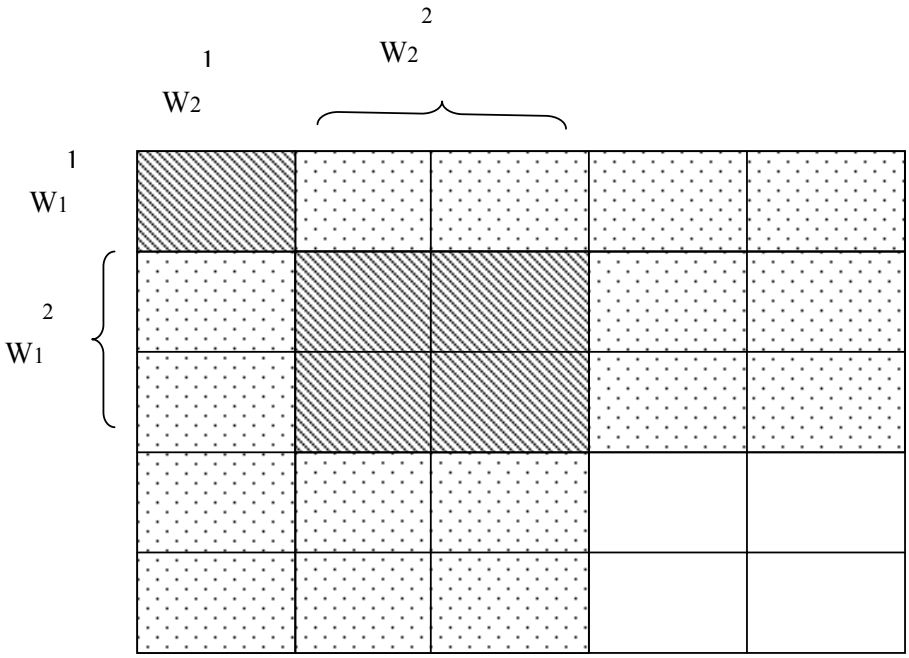


Figure 1: Relevant areas of steps 2-4

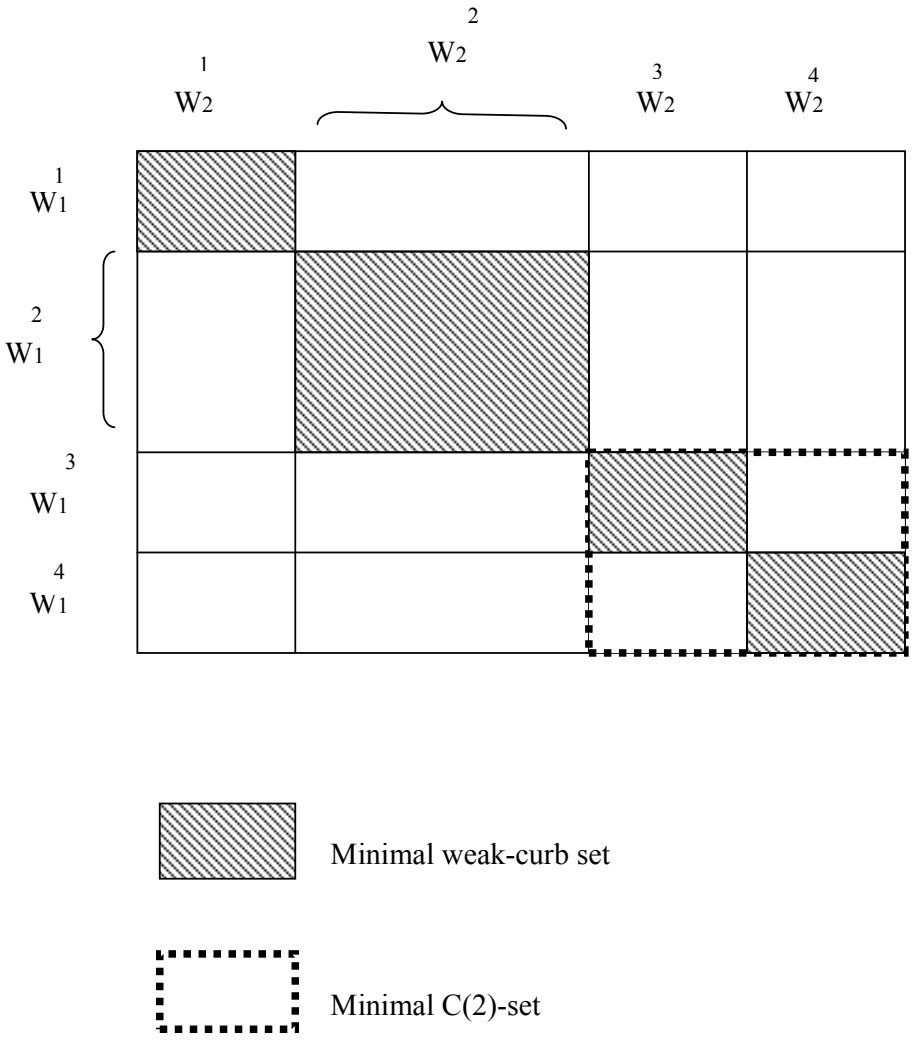

1
W2
2
W2
3
W2
4
W2
1
W1
2
W1
3
W1
4
W1
Minimal weak-curb set
Minimal C(2)-set

Figure 2

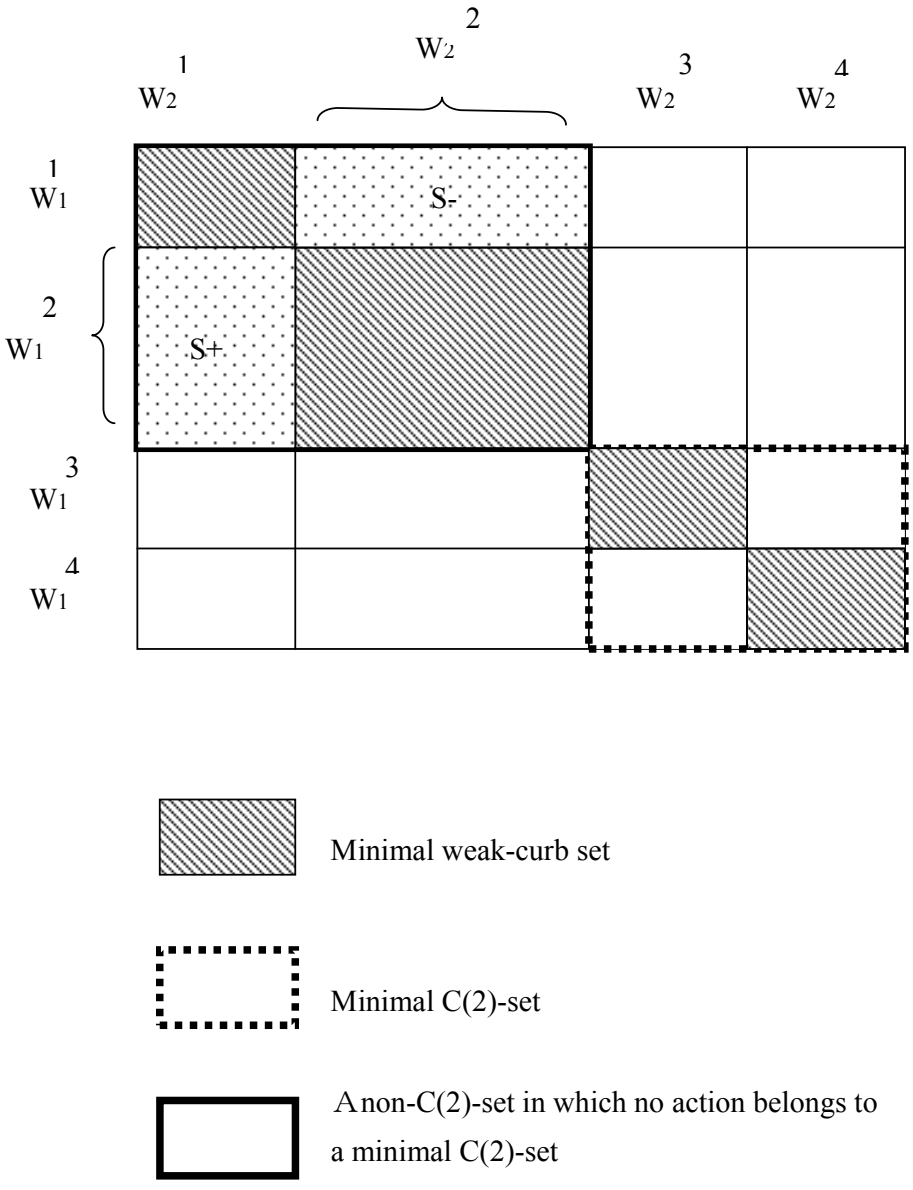
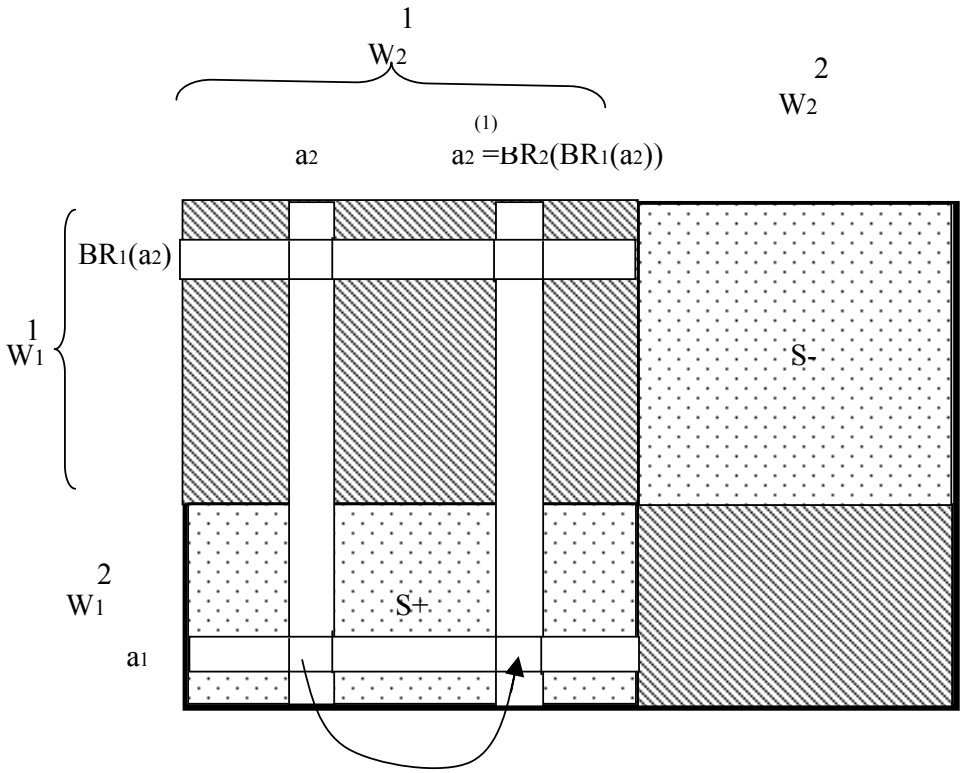

2
W2
1
W2
3
W2
4
W2
1
W1
S-
2
W1
S+
3
W1
4
W1
Minimal weak-curb set
Minimal C(2)-set
A non-C(2)-set in which no action belongs to a minimal C(2)-set

Figure 3



Minimal weak-curb set

A non-C(2)-set in which no action belongs to a minimal C(2)-set

Figure 4