Learning, Experimentation, and Long-Run Behavior in Games*

Andreas Blume

Department of Economics

University of Iowa

November 8, 1994

Abstract

This paper investigates a class of population-learning dynamics. In every period agents either adopt a best reply to the current distribution of actual play, or a best reply to a sample, taken with replacement, from the distribution of intended play (the strategies adopted at the end of last period), or they are inactive. If sampling with replacement and being inactive have strictly positive probability, these dynamics converge globally to *minimal curb sets* in the absence of mistakes. For two-player $i \times j$-games, $i, j \leq 3$, the same result holds even if only best responding to actual play and being inactive have positive probability. If players make mistakes in the implementation of their strategies, these dynamics select among *minimal curb sets*.

*I wrote the first draft of this paper while I enjoyed the hospitality of CentER. I thank, without implicating, Sjaak Hurkens and Eric van Damme for comments. I have also benefitted from comments made in the IOMEGT workshop at the University of Iowa. I am grateful to the NSF for financial support.

1 Introduction

This paper characterizes the long-run outcomes of a class of learning dynamics in games. The characterization is in terms of properties of subsets of the space of strategy profiles of the underlying games. Young [1993] has shown that under some conditions on players' memory and the completeness of their information, *adaptive play* converges almost surely to a pure strategy equilibrium, provided the game satisfies an acyclicity requirement. The present paper is in the same spirit. It looks at a different dynamic and, more importantly, it drops the acyclicity condition. Without this condition, the question arises which objects can take the place of the pure strategy Nash equilibria. In general, one suspects that this will depend on details of the dynamic. This paper argues that there are interesting classes of dynamics whose long-run outcomes can be characterized in terms of *curb (closed under rational behavior)* sets.

A product set of strategies is closed under inclusion of best replies if it contains all best responses to independent beliefs supported on itself. Basu and Weibull [1991] who first examined the properties of such sets, refer to them as *closed under rational behavior (curb).* Curb sets which do not properly contain another curb set are referred to as *minimal curb sets.* In generic normal form games these coincide with persistent sets (the extreme points of persistent retracts), Kalai and Samet [1984], Balkenborg [1992]. While the set of rationalizable strategies, Bernheim [1984] and Pearce [1984], is a maximal fixed point under the best reply mapping, a minimal curb set is a minimal fixed point under this mapping.

Curb sets have a number of attractive features. They share with strict equilibria the property that they contain all best replies against themselves. Every curb set contains the support of a Nash equilibrium; such equilibria are referred to as curb equilibria. Blume [1994] shows that in games with one-sided pre-play communication, the minimal curb condition selects the communicating player's favorite equilibrium if it is not too risky. Hurkens [1993] demonstrates that in pre-play communication games where messages are made distinct via nominal message costs, all minimal curb equilibria are efficient for the communicating players, if the underlying game has a strict equilibrium which gives the communicating players their most preferred payoff.

Like strict equilibria, curb sets will be locally stable under a large class of plausible dynamic adjustment rules. The question I want to pose in this paper is whether it is possible to provide a firmer dynamic foundation for minimal curb sets. This requires that one address two issues. Are there dynamics which converge globally to minimal curb sets? Are there dynamics which select among minimal curb sets?

Hurkens [1994] provides one answer to the first question. He examines a dynamic in the spirit of Young [1993] which converges almost surely to a minimal curb set from any initial condition. In his dynamic, only one pair of agents plays in any given time period. Each of them takes a possibly incomplete sample from finite length histories of past play and best responds to *some* distribution over the sample. The fact that only the support of the sample matters, guarantees that every belief over strategies in the current state is possible. Therefore the dynamic will eventually leave any set of strategies that is not curb. Finite length histories guarantee that once the process has spent sufficient time in a minimal curb set it cannot exit the minimal curb set anymore. Hurkens shows that his dynamic converges globally to minimal curb sets, almost surely. He then goes on to ask whether adding mutations to his dynamic yields selection among minimal curb sets. This is not the case because as soon as a mistake enters the current state, it is possible that the active players attribute *any* probability to the corresponding strategy. Therefore one mistake is sufficient to upset any minimal curb set in Hurkens' framework.

I want to propose a different dynamic. I consider large populations of agents. All agents play in every period; when they play, they use the strategies they had *adopted* at the end of last period, unless they are experimenting. If they are experimenting, they randomize, choosing each of their available strategies with strictly positive probability. Agents differ in how they process information. In each period every agent either *best responds*, or *gathers information* or is *inactive.* Best responding agents learn the true current distribution of play from playing against the entire population. Agents who gather information sample (with replacement) from the distribution of strategies agents had adopted at the end of last period. Either of these active agent groups adopt a best reply against their information.[1] Inactive agents carry over the strategy they had adopted at the end of last period into the next period. Because play partners are identifiable, the information available to best responders is modelled as coming from sampling without replacement and for simplicity as learning the true distribution of play, whereas information gathering is potentially indirect and therefore modelled as sampling with replacement.

I derive two main results; one under the condition that the mistake probability is zero and one for positive but small mistake probabilities. Without mistakes the dynamic converges to minimal curb sets regardless of the initial condition. With mistakes the dynamic selects among minimal curb sets. For two-player games with two exhaustive minimal curb sets, I characterize the condition for selection of one of the minimal curb sets; this condition reduces to Harsanyi-Selten risk dominance in 2×2 games. Intuitively, the two sets of states where the populations play entirely according to one of the two minimal curb sets *(minimal curb states sets)* are exceptional. The dynamic can leave these sets of states only if sufficiently many mistakes occur simultaneously. Any other state, outside of these sets, belongs to the basins of attraction of *both* minimal curb state sets. Thus all that matters is how many mistakes it takes to upset either one of the two minimal curb state sets.

I also investigate to what extent we need sampling with replacement or as in Hurkens [1994] a direct assumption that every distribution which is supported on the current state has positive probability. I show that we may be able to do without such assumptions. At least in two-player $i \times j$-games with $i, j \leq 3$ a simple best-reply rule where agents are either inactive or move to one of their best replies, converges globally to minimal curb sets almost surely.

Besides the problem of finding dynamics which lead to and select among minimal curb sets in a game, there is the dual problem of when it is possible to find simple characterizations of stable sets of a dynamic in terms of the game. For a Markov process with stationary transition probabilities the stable sets are the recurrent communication

[1] For technical reasons I will have to work with almost best replies or restrict the analysis to a generic class of games.

classes of the process. As Young [1993] has shown, one can select among the recurrent communication classes by considering limits of perturbed processes as the perturbation vanishes. The limiting distribution will have its support concentrated on the communication classes with the least stochastic potential. In general, it will be difficult to characterize these communication classes in terms of the underlying game. The present paper proposes a class of dynamics for which such a characterization is possible; for these dynamics the recurrent communication classes of the unperturbed dynamic correspond to the minimal curb sets of the underlying game. The correspondence is as follows: for a given recurrent communication class there is exactly one minimal curb set such that each state has support only on this minimal curb set; conversely, given any minimal curb set, every state with support on this minimal curb set belongs to one and the same recurrent communication class. Furthermore, for a non-trivial class of games one can derive sufficient conditions for the selection of a particular minimal curb set via perturbed dynamics. Interestingly, it turns out, that the payoffs in the equilibria belonging to the minimal curb sets play a secondary role as far as selection is concerned. Suppose one starts with a strict equilibrium that is selected by the perturbed dynamic. If one then replaces this equilibrium with a game that is a minimal curb set in the newly formed game and has a unique equilibrium with the same payoffs as the original equilibrium, this minimal curb set need not be selected by the perturbed dynamic.

The paper is organized as follows. The next section describes the model. Section 3 introduces the dynamic without mutations and derives the global convergence to minimal curb sets. Section 4 introduces mistakes in the implementation of strategies and demonstrates that the dynamic selects among exhaustive minimal curb sets in two-player games. Section 5 concludes.

2 The Setup

Consider a finite set of populations P with typical element $p \in P$. Denote the size of population p by N_p. Each population corresponds to one of the players in the game $G = \{S, u\}$. S_p is the finite set of pure strategies available to a type p player; assume that $N_p > \#(S_p)$, the cardinality of player p's strategy space. $S := \times_{p \in P} S_p$ with typical element $s \in S$, and u_p is a type p player's utility function; $u_p : S \rightarrow \Re$. Let Σ_p denote a type p player's set of mixed strategies, and $\Sigma := \times_{p \in P} \Sigma_p$. A typical element of Σ will be denoted by σ. If we exclude the pth element from σ, the resulting vector will be written as σ_{-p}. u_p extends to Σ in the usual way.

For any finite set X, let $\Delta(X)$ stand for the set of probability distributions over X. Let $BR_p(\cdot)$ denote player p's pure best reply correspondence, and define $BR(\sigma) := \times_{p \in P} BR_p(\sigma_{-p})$. I will also use the natural extension of $BR(\cdot)$ to sets of strategies as arguments.

Basu and Weibull [1991] introduced the notion of *curb (closed under rational behavior)* sets. A product set of strategies $Q = \times_{p \in P} Q_p$, $Q_p \subset S_p$, is closed under inclusion of best replies (curb) if each Q_p is nonempty and

$$BR(\times_{p \in P} \Delta(Q_p)) \subseteq Q.$$

If a curb sets does not properly contain another curb set, it is called minimal. The strategies which form a minimal curb set are called curb strategies, and equilibria belonging to minimal curb sets are curb equilibria.

Blume [1994] and Hurkens [1993] show that the curb equilibrium requirement selects efficient outcomes in games with pre-play communication. Blume considers games with costless messages; Hurkens analyzes the case of nominal message costs. Consider the two games below. Both games have two minimal curb sets corresponding to the two strict Nash equilibria (U, L) and (D, R).

	L	R
U	3,3	0,0
D	0,0	1,1

$\mathbf{G_1}$

	L	R
U	9,9	0,8
D	8,0	7,7

$\mathbf{G_2}$

If we allow player one to send one of two messages m_1 or m_2 before playing the game, the reduced normal forms corresponding to the G_1 and G_2 are.

	LL	LR	RL	RR
(m_1, U)	3,3	3,3	0,0	0,0
(m_1, D)	0,0	0,0	1,1	1,1
(m_2, U)	3,3	0,0	3,3	0,0
(m_2, D)	0,0	1,1	0,0	1,1

$\mathbf{\Gamma_1}$

	LL	LR	RL	RR
(m_1, U)	9,9	9,9	0,8	0,8
(m_1, D)	8,0	8,0	7,7	7,7
(m_2, U)	9,9	0,8	9,9	0,8
(m_2, D)	8,0	7,7	8,0	7,7

$\mathbf{\Gamma_2}$

If messages are costless, then Blume [1994] shows that all curb equilibria in Γ_1 support the efficient payoff pair $(3,3)$. This result generalizes provided a condition that trades off the risk of the efficient equilibrium in the underlying against the size of the message space is satisfied. In Γ_2 which is based on G_2, all equilibria are curb equilibria because there is a tension between risk dominance and Pareto dominance in the underlying game. If messages carry a nominal cost which distinguishes them, these results can be strengthened considerably. Hurkens [1993] shows that with nominal message costs, $0 \leq m_1 < m_2$, $\{(m_1, U)\} \times \{LL, LR\}$ is the unique minimal curb set in both of the above communication games. He shows that this result generalizes to n-player games in which a subset of the player set can send a message.

3 Dynamics

In this section I describe the learning dynamic, and characterize its long-run outcomes in the absence of experimentation. I will show that the process converges almost surely to a curb set, regardless of the initial population state. In the following section I will examine this process further under the condition that the experimentation probability is different from zero.

The *state* of population p at time t is given by the vector $\omega_{p,t} = \{s_{it}\}_{i \in p}$. The state of the dynamic system at time t is given by $\omega_t = \{\omega_{p,t}\}_{p \in P}$. In period t state ω_{t-1} is replaced by ω_t according to the following rule. Agent i enters period t with pure strategy $s_{i,t-1}$.

Every agent in p plays against all agents in $P\backslash p$. When she plays in period t, an agent uses strategy $s_{i,t-1}$, unless she experiments (or makes a mistake). Each agent experiments with probability $\eta \geq 0$; in that case she chooses each of her pure strategies with strictly positive probability. There are three different ways in which agents process information in period t and thereby generate the new state s_t. With probability λ_1 an agent adopts a best reply against the current distribution of actual play in period t, with probability λ_2 she gathers information about the strategies that were adopted last period, and with probability $\lambda_3 = 1 - \lambda_1 - \lambda_2$ she is inactive in period t. Subsequently I will refer to agents in these various roles as, best responding, information gathering and inactive players. A best responding agent meets all agents from populations she does not belong to and learns the true distribution of current play (including mistakes) in the current period in those populations; she then adopts a best reply against this distribution. An information gathering agent takes a sample (with replacement) from the strategies that were adopted in period $t-1$ (the intended play of period t); she then adopts an ϵ-best reply against uncorrelated beliefs based on this sample. An agent who is inactive in period t passes through that period without changing her strategy.

Play with mistakes in period t generates a temporary state $\tilde{\omega}_t$ to which best responding players best respond at the end of period t. Note that for each agent $i \in p$ each partial temporary state $\tilde{\omega}_{-p,t}$ naturally can be identified with an uncorrelated belief $\mu_{-p} = \mu_1 \times \cdots \times \mu_{p-1} \times \mu_{p+1} \times \cdots \times \mu_{\#(P)}$ for agent $i \in p$ over $S_{-p} := \times_{q \neq p} S_q$, which itself can be identified with an element of $\Sigma_{-p} = \times_{q \neq p} \Sigma_q$. This belief is based on the observed relative frequencies of strategies in the populations not including i. With this identification of states and beliefs we can say that a best responding agent $i \in p$ adopts a pure strategy in $BR_i(\tilde{\omega}_{-p})$. For the remainder of this section I will set the experimentation probability to zero; I return to the case of $\eta > 0$ in the next section.

Denote the period t sample from population p obtained by an information gathering agent $i \notin p$ by X_{ipt}. Agent i's entire period t sample is then $X_{it} := \{X_{ipt}\}_{p \in P, p \not\ni i}$ Like states, samples give rise to uncorrelated beliefs. Therefore it makes sense to consider $BR_i^{\epsilon}(X_{it})$, the set of agent i's pure ϵ-best replies against uncorrelated beliefs based on the sample X_{it}. An information gathering agent i adopts one of these ϵ-best replies;[2] in case of indifference, she randomize, putting strictly positive probability on each of the strategies in $BR_i^\epsilon(X_{it})$. Let sample sizes be time invariant and denote the size of player i's sample from population p, $i \notin p$, by N_{ip}.

The dynamic process described here is a Markov chain with stationary transition probabilities on the state space, denoted by Ω. The first objective of this paper is to characterize the recurrent communication classes of this Markov chain. The recurrent communication classes are subsets of Ω such that (i) from every state there is a finite length sequence of positive probability transitions to at least one of these classes, (ii) within each class every state can be reached from every other state via a finite length sequence of positive probability transitions, and (iii) no state outside one of the classes can be reached from a state inside through a positive probability transition. Since (minimal) curb sets will figure prominently in this characterization, define a set of states supported entirely on one (minimal) curb set, and including all such states as a (minimal-)curb-state set.

Let

$$\Psi(\Theta) := \{\sigma \in \Sigma | \text{supp}(\sigma_i) \subseteq \Theta_i\} \ \ \forall \Theta \subseteq S.$$

For any subset Θ of the set of pure strategy profiles, this is the set of all mixed strategy profiles or equivalently uncorrelated beliefs with support in Θ. For any such Θ one can define the set

$$V(\Theta) := \Theta \cup BR(\Psi(\Theta)) \ \ \forall \Theta \subseteq S$$

consisting of the union of Θ and all best replies against uncorrelated beliefs concentrated on Θ. Let V^t denote the t-fold iteration of V. It is easily seen that, starting with a set Θ

[2] I use ϵ-best replies because the sampling process only allows one to approximate the set of possible beliefs over a given support of strategies. Alternatively one could proceed like Hurkens and argue directly in terms of supports; i.e. one could simply postulate an updating rule in which players can move to any strategy which is a best reply to some beliefs over a given support. Also, in a generic class of games, we can replace ϵ-best replies by best replies in our dynamic. The generic property to look for is that any strategy that is a weak best reply to some beliefs concentrated on support Q, is a strict best reply against some beliefs concentrated on Q as well.

one reaches a fixed point of V by iterating V sufficiently often.

Lemma 1 $\forall \Theta \subseteq S,\ \exists T : (\forall t > T, V^{t+1}(\Theta) = V^t(\Theta))$.

Proof: $V : 2^S \to 2^S$ is monotonic and 2^S is finite. □

We next need notation to describe the fixed point that is reached if one starts the iteration with the support $\Theta(\sigma)$ of a strategy profile σ.

Let

$$\begin{aligned} \Theta(\sigma) &:= \{s \in S | s_i \in \mathrm{supp}(\sigma_i)\} \\ t(\sigma) &:= \min\{t \in \mathbf{N} | V^{t+1}(\Theta(\sigma)) = V^t(\Theta(\sigma))\} \\ W(\sigma) &:= V^{t(\sigma)}(\Theta(\sigma)) \end{aligned}$$

$t(\sigma)$ is the minimal number of periods needed before one reaches the fixed point from $\Theta(\sigma)$, and $W(\sigma)$ is the fixed point reached from σ.

Lemma 2 *$W(\sigma)$ contains a minimal curb set for all $\sigma \in \Sigma$.*

Proof: $W(\sigma)$ is closed under inclusion of best replies. □

The set of states Ω can be identified with a finite subset of the set of mixed strategies. The dynamics can be described by a transition probability $\phi(\cdot|\cdot)$ such that $\forall\ \sigma,\ \tau \in \Omega$, $\phi(\tau|\sigma)$ denotes the probability that the system will be in state τ in period $t+1$, if in period t it is in state σ. ϕ depends on population sizes, sample sizes and ϵ. Let $\mathcal{N} := \{\{N_p\}_{p\in P}, \{N_{ip}\}_{p\in P, i\notin p}\}$. Let $\phi_{\epsilon,\mathcal{N}}$ be the transition probability as a function of ϵ and $\mathcal{N}$.

The set Ω does not contain states corresponding to every belief. Therefore it is possible that our dynamic with best replies, instead of ϵ-best replies, does not leave a given set of states even though that set is not a curb-state set. The following game provides an example

	t_1	t_2	t_3
s_1	1,1	1,0	0,b
s_2	1,0	1,a	0,b
s_3	0,0	0,0	.1,.1

where $a = \sqrt{2}$, and $b = \frac{\sqrt{2}}{1+\sqrt{2}}$.

In this game only the strict equilibrium (s_3, t_3) forms a minimal curb set. However the product set $\{s_1, s_2\} \times \{t_1, t_2\}$ is closed under inclusion of best replies against beliefs for which each probability must be a rational number. Only if the column player puts probability b on the first strategy of her opponent is her third strategy a best reply against beliefs concentrated on the first two strategies of her opponent.

On the other hand, with sufficiently large sample sizes any belief can be approximated arbitrarily closely. Therefore, with ϵ-best replies there is a chance that our dynamic eventually leaves every set that is not a curb-state set.[3] This motivates the next lemma which says that every best reply to a given belief is an ϵ-best reply to an open neighborhood of that belief. Thus, if we can approximate beliefs arbitrarily closely, the set of best replies to any product set of strategies is a subset of the set of ϵ-best replies to the finite approximation of the same set, provided the approximation is sufficiently close.

Lemma 3 $\forall \epsilon > 0,\ \forall \sigma \in \Sigma,\ \forall s_i \in BR_i(\sigma), \exists \delta > 0 :\ |\tilde{\sigma} - \sigma| < \delta\ \Rightarrow s_i \in BR_i^{\epsilon}(\tilde{\sigma})$.

Proof: Suppose not, and let $s_i \in BR_i(\sigma)$. Then there exists $\tilde{\sigma}_n \to \sigma$, $t_i(\tilde{\sigma}_n) \in S_i$ such that

$$u_i(t_i(\tilde{\sigma}_n), \tilde{\sigma}_n) > u_i(s_i, \tilde{\sigma}_n) + \epsilon \ \ \forall n.$$

Compactness of the strategy space and continuity of u imply that there exists a $t_i \in S_i$ such that

$$u_i(t_i, \sigma) \geq u_i(s_i, \sigma) + \epsilon,$$

[3] In a generic class of games this problem does not arise.

which contradicts $s_i \in BR_i(\sigma)$. □

For any given $\sigma \in \Omega \subset \Sigma$, $\phi_{\epsilon,\mathcal{N}}$ induces a probability distribution over the set 2^S of supports. This is the probability that next period's state will have a certain support given that the current state is σ. Let the probability of support $\Theta \in 2^S$ given the current state is σ be denoted $P_{\epsilon,\bar{N}}(\Theta|\sigma)$, given the transition probability $\phi_{\epsilon,\mathcal{N}}(\cdot|\cdot)$. Let $P^t_{\epsilon,\mathcal{N}}(\Theta|\sigma)$ denote the probability of support Θ after t periods if the initial state is σ. In particular $P^1_{\epsilon,\mathcal{N}}(\Theta|\sigma) = P_{\epsilon,\mathcal{N}}(\Theta|\sigma)$. Let $N_{\min} := \min\{N_{ip}\}_{i \in I, p \in P}$.

A central characteristic of our learning rule is that from any state σ there is positive probability that next period's state will have support $V(\Theta(\sigma))$. Iterating this argument, one may conclude that from any σ there is positive probability that after a finite number of periods the state has a support which is a fixed point of $V(\cdot)$. This is the content of the following lemma.

Lemma 4 *If* $\lambda_2, \lambda_3 > 0$, *then* $\forall \epsilon > 0$, $\exists \bar{N}, T \; : \; N_{\min} > \bar{N} \Rightarrow P^T_{\epsilon,\mathcal{N}}(W(\sigma)|\sigma) > 0$, $\forall \sigma \in \Omega$.

Proof: Let $\Omega(\sigma)$ be the set of possible samples with replacement, given the current state σ.. Given the identification of samples with beliefs, $\Omega(\sigma)$ is a finite approximation of $\Psi(\Theta(\sigma))$. Furthermore, $\Omega(\sigma)$ converges to $\Psi(\Theta(\sigma))$ in the Hausdorff sense as $N_{\min} \to \infty$. Let $BR_i(\Psi(\Theta(\sigma)))$ be the (finite) set of (pure) best replies by agent i to the beliefs concentrated on $\Theta(\sigma)$. For every i, $\forall s_i \in BR_i(\Psi(\Theta(\sigma)))$, choose $\tau(s_i) \in \Psi(\Theta(\sigma))$ such that $s_i \in BR_i(\tau(s_i))$. Note that for every $\delta > 0$, $\exists N_\delta : N_{\min} > N_\delta \Rightarrow \forall \sigma, \forall i, \forall s_i \in BR_i(\Psi(\Theta(\sigma))), \exists \tilde{\tau}(s_i) \in \Omega(\sigma)$ such that $|\tau(s_i) - \tilde{\tau}(s_i)| < \delta$. To see this note that there are finitely many i, finitely many combinations of $\Theta(\sigma)$ and $s_i \in BR_i(\Psi(\Theta(\sigma)))$, and that each single $\tau(s_i)$ can be approximated by a belief in $\Omega(\sigma)$.

By Lemma 3 $\forall \epsilon$, $\forall \tau(s_i), \exists \delta > 0 : \; |\tau(s_i) - \tau| < \delta \Rightarrow s_i \in BR^\epsilon_i(\tau)$. Since there are finitely many such $\tau(s_i)$ to consider across all individuals and all supports, we can interchange quantifiers to obtain $\forall \epsilon$, $\exists \delta : \forall i, \forall \tau(s_i), |\tau(s_i) - \tau| < \delta \Rightarrow s_i \in BR^\epsilon_i(\tau)$.

Combining the last two observations, we may conclude that: $\forall \epsilon > 0$, $\exists \bar{N} : N_{\min} > \bar{N} \Rightarrow \forall \sigma, \forall i, \forall s_i \in BR_i(\Psi(\Theta(\sigma)))$, $\exists \tilde{\tau}(s_i) \in \Omega(\sigma)$ such that $s_i \in BR^\epsilon_i(\tilde{\tau}(s_i))$. Since all samples from $\Omega(\sigma(t))$ have positive probability, each $s_i \in BR_i(\Psi(\Theta(\sigma(t))))$ has positive probability of being in the support of $\sigma(t+1)$; because $\lambda_3 > 0$, any $s_i \in \Theta(\sigma(t))$, also has positive probability of being in the support of $\sigma(t+1)$.

Since $N_p > \#(S_p)$, with positive probability all the strategies in $V(\Theta(\sigma(t)))$ are present in the population in period $t+1$. Therefore, $\forall \epsilon > 0, \exists \bar{N} : N_{\min} > \bar{N} \Rightarrow P_{\epsilon,\mathcal{N}}(V(\Theta(\sigma))|\sigma) > 0$, $\forall \sigma \in \Omega$. The conclusion follows by applying this last observation repeatedly and combining it with Lemma 1. □

According to the lemma there exists an upper bound on the minimal number of positive probability transitions it takes from any initial state to reach a state which "covers" a curb set. At that point there exists a positive probability transition into a minimal curb set.

Corollary 1 *If $\lambda_2, \lambda_3 > 0$, then $\forall \epsilon > 0$, $\exists \bar{N}$, T' such that for $N_{\min} > \bar{N}$, from any initial state $\sigma \in \Omega$, the system moves into a minimal-curb-state set after no more than T' iterations with positive probability.*

Proof: From the proposition, after T steps the system reaches a state whose support "includes" a minimal curb set. There is positive probability that in the next round all agents are active and draw samples from the curb set. Let $T' = T + 1$. □

The following lemma verifies that if, for a given game, ϵ is chosen sufficiently small, then the learning dynamic cannot exit a minimal-curb-state curb set once it has entered it.

Lemma 5 $\exists \bar{\epsilon} > 0 : \forall 0 < \epsilon < \bar{\epsilon}$, *if $\sigma(t)$ is an element of a curb-state set Θ, then* $\operatorname{supp}(\sigma(t+k)) \subseteq \Theta$, $\forall k \geq 0$

Proof: If $s_i \notin BR_i(\Theta)$, then there exists $\bar{\epsilon}(s_i) > 0$ such that $\forall 0 < \epsilon < \bar{\epsilon}(s_i)$, $s_i \notin BR_i^{\epsilon}(\Theta)$. Consider $\bar{\epsilon} := \min\{\bar{\epsilon}(s_i) | i \in I, s_i \notin BR_i(\Theta), \Theta \subseteq S\}$. □

We are now ready to state the main result of this section. If players use almost best replies and sample sizes are sufficiently large, then regardless of the initial conditions, the learning process will eventually end up in one of the minimal-curb-state sets.

Proposition 1 *If $\lambda_2, \lambda_3 > 0$, then, $\exists \bar{\epsilon} > 0$, such that $\forall 0 < \epsilon < \bar{\epsilon}$, $\exists \bar{N}$ such that for $N_{\min} > \bar{N}$, and for any initial state $\sigma \in \Omega$, the learning process converges almost surely to a minimal-curb-state set.*

Proof: For a given ϵ let $N_{\min}$ and T' be given as in Corollary 1 such that

$$P_{\epsilon,\mathcal{N}}(\sigma(t+T') \in \text{minimal-curb-state set}|\sigma(t) = \sigma) \geq \pi > 0 \;\; \forall \sigma \in \Omega.$$

Then

$$P_{\epsilon,\mathcal{N}}(\sigma(t+kT') \notin \text{minimal-curb-state set}|\sigma(t) = \sigma) \leq (1-\pi)^k.$$

Thus the probability that the system does not converge to a minimal-curb-state set equals

$$\lim_{k\to\infty} (1-\pi)^k = 0.$$

□

Not only does the learning dynamic converge globally almost surely to one of the minimal-curb-state sets. Inside such sets every state is reached from every other state via a finite length sequence of positive probability transitions. This follows from lemma 4. Therefore we have the following corollary:

Corollary 2 *If $\lambda_2, \lambda_3 > 0$, then, $\exists \bar{\epsilon} > 0$, such that $\forall 0 < \epsilon < \bar{\epsilon}$, $\exists \bar{N}$ such that for $N_{\min} > \bar{N}$, the minimal-curb-state sets are the recurrent communication classes of the learning dynamic.*

3.1 A Simple Best-Reply Rule

In some interesting classes of games one obtains convergence to minimal-curb-state sets from a simple best-reply rule.[4] Consider a dynamic in which agents are either inactive,

[4] Again, in general we have to consider ϵ- best replies, and large populations. I will largely ignore these details as it is easy for the reader to fill them in where needed.

with probability $1-\mu$, or move to a best reply, with probability μ, $0<\mu<1$. One can easily check that this process converges almost surely to the unique minimal-curb-state set in the pre-play communication game with reduced normal form Γ_1. In this case there is no need to introduce sampling with replacement to generate a rich enough set of beliefs out of the current state.

It is easy to see that this observation generalizes for this class of communication games. However, it is not clear how far one can extend it beyond this class. In this subsection I will show that the observation is valid for all two-player $i \times j$-games with $i, j \leq 3$. In this class of games it is sufficient that some agents move to a best reply while others don't to generate beliefs which will induce exit from any product set, that does not contain a minimal curb set.

For technical reasons I will again replace best replies in the dynamic by ϵ-best replies and refer to the "ϵ-best-reply rule." In the proofs I will argue in terms of best replies. This suffices for generic games; the arguments for general games in terms of ϵ-best replies are analogous to the ones made above and therefore omitted. There are two populations, p=1,2.

Proposition 2 *Let G be any $i \times j$-game with $i, j \leq 3$. Then, for all $0<\mu<1$, there exists $\bar{\epsilon}$, such that for all $0<\epsilon<\bar{\epsilon}$ there exists $\bar{N}$ such that $N_p > \bar{N}$ implies that under the ϵ-best-reply rule the process will almost surely converge to a minimal-curb-state set.*

Proof: Note that from any state there is positive probability that in one step the process moves to a state such that within each population all agents use the same strategy. If one of these strategies is a minimal curb strategy, we are done, because there is positive probability that only the other population moves and that within that population every agent moves to a best reply. Therefore continue under the assumption that we start with a state in which every member of a population uses the same strategy which is not a minimal curb strategy.

Since this initial state (s_1, t_1) is not curb, the dynamic process exits this state with positive probability, and moves to a state in which one population is concentrated on one and the other is concentrated on two strategies; e.g., the new state is supported on $\{s_1, s_2\} \times \{t_1\}$. If s_2 is a curb strategy we are done because (s_2, t_1) was one of the states reached with positive probability from (s_1, t_1). Otherwise, $\{s_1, s_2\} \times \{t_1\}$ is not curb, which means that either s_3 is a best reply to t_1, in which case we are done because there are only three strategies and one of them has to be curb, or there exist beliefs concentrated on $\{s_1, s_2\}$ such that the column player has a best reply which is not t_1; without loss of generality let it be t_2. If t_3 is also a best reply to some beliefs on $\{s_1, s_2\}$ we are done because either t_2 or t_3 would have to be a curb strategy. Suppose not, i.e. let t_2 be the only best reply and let it not be a curb strategy. Note that from the initial state (s_1, t_1) (almost) every distribution over the two strategies s_1 and s_2 has positive probability. Therefore, with positive probability we move to a state supported on $\{s_1, s_2\} \times \{t_1, t_2\}$. In particular, for (almost) every distribution over $\{t_1, t_2\}$, there is a corresponding state which can be reached with positive probability.

None of the strategies s_1, s_2, t_1, and t_2 is a curb strategy. And since there is no belief over $\{s_1, s_2\}$ to which t_3 is a best reply, it must be the case that s_3 is a best reply to some beliefs over $\{t_1, t_2\}$. Suppose we are at a positive probability state where the distribution over $\{t_1, t_2\}$ is such that s_3 is a best reply. At that point there is positive probability that the entire row population moves to s_3 which from the foregoing must be a curb strategy. □

There is still the question of whether it is possible that the dynamics may end up being confined to a subset of a minimal-curb-state set. Ideally we would want to show that the recurrent communication classes of the ϵ-best reply dynamic coincide with the minimal-curb-state sets. I will prove a somewhat weaker result.

Proposition 3 *Let G be any two-player $i \times j$-game with $i, j \leq 3$. Then, for all $0 < \mu < 1$, there exists $\bar{\epsilon}$, such that for all $0 < \epsilon < \bar{\epsilon}$ there exists $\bar{N}$ such that for $N_p > \bar{N}$ the following holds: For every minimal curb set $Q = Q_1 \times Q_2$ of G, every $q \in Q_i$, and every recurrent communication class C supported on Q, there is at least one state in C in which q has positive weight.*

Proof: Obvious for $1 \times j, j = 1, 2, 3$, minimal curb sets.

What about 2×2? Consider a minimal curb set of the form $\{s_1, s_2\} \times \{t_1, t_2\}$. Without loss of generality we can start with a pure strategy combination, say (s_1, t_1). Also wlog s_2 is a best reply, and there exists a belief, β, over $\{s_1, s_2\}$ to which t_2 is a best reply. The result for 2×2 minimal curb sets follows because we can move any fraction of the population to a best reply.

Next consider minimal curb sets of the form $\{s_1, s_2\} \times \{t_1, t_2, t_3\}$. Without loss of generality we can start the process at (s_1, t_1). Then either s_2, or t_2, or t_3 is a best reply.

Suppose first that s_2 is a best reply to t_1. Then the dynamic can generate all possible beliefs of the column player over $\{s_1, s_2\}$. Since we are dealing with a minimal curb set, there must exist beliefs β_2 and β_3 concentrated on this set such that $t_j \in BR(\beta_j), j = 2, 3$.

Now suppose instead that s_2 is not a best reply to t_1. Then, if t_2 and t_3 are both best replies to s_1, the dynamic can generate all distributions over $\{t_1, t_2, t_3\}$ and there will be at least one distribution over these three strategies which makes s_2 a best reply.

It remains to consider the case where only t_2 is a best reply to s_1. In that case s_2 must be a best reply to t_2, for otherwise $\{s_1\} \times \{t_1, t_2\}$ would form a minimal curb set. This is analogous to the case where s_2 was a best reply to t_1.

Next consider minimal curb sets of the form $\{s_1, s_2, s_3\} \times \{t_1, t_2, t_3\}$. As before, wlog, we can start the dynamic out at a state corresponding to the pure strategy combination (s_1, t_1). Also, wlog, s_2 is a best reply to t_1, which means that we can move to any belief concentrated on $\{s_1, s_2\}$.

If s_3 is a best reply as well, we are done because we can move to any mixture over $\{s_1, s_2, s_3\}$ and there is at least one such mixture for each t_2 and t_3 which makes them best replies.

Therefore suppose that s_3 is not a best reply to t_1. If there are beliefs β_2 and β_3 over $\{s_1, s_2\}$ such that $t_j \in BR(\beta_j), j = 2, 3$, we can generate all beliefs over $\{t_1, t_2, t_3\}$ in the following way. First move to a state supported on $\{s_1, s_2\} \times \{t_1\}$ corresponding to the belief $\beta_j, j = 2, 3$ with the least weight on s_2, say β_2. Then, simultaneously move the row population to β_3 and the desired fraction of the column population to t_2. In the next step move the desired fraction of the column population to t_3. Again we are done because we have been able to generate all possible beliefs over $\{t_1, t_2, t_3\}$

Suppose next that only for t_2 there is a belief β_2 supported on $\{s_1, s_2\} \times \{t_1\}$ such that $t_2 \in BR(\beta_2)$. Then s_3 must be a best reply to some beliefs over $\{t_1, t_2\}$. By an argument analogous to the one just given it follows that we can generate all beliefs over $\{s_2, s_3\}$. If there is such a belief such that t_3 is a best reply, we are done. Otherwise, s_1 must be a best reply to some beliefs concentrated on t_1, t_2. Note also that t_2 must be a best reply to either s_1 or s_2. In either case we can repeat the construction from the previous paragraph to generate all beliefs over $\{s_1, s_2, s_3\}$ which concludes the argument. □

4 Selection

In the previous section I considered learning without experimentation, mistakes or mutations. I showed that from any initial condition the dynamic converges almost surely to one of the minimal-curb-state sets. The unperturbed dynamic does not select among curb sets. Work by Young [1993], Kandori, Mailath and Rob [1993], Ellison [1993] shows that similar dynamics select among strict Nash equilibria, provided they are augmented to allow for mistakes. Samuelson [1993], Nöldeke and Samuelson [1993] [1994] investigate selection among nonsingleton recurring communication classes. Hurkens [1994] shows that an intuitive class of dynamics which converges globally to curb sets does not select among them once mistakes are added. I will show in this section that adding mistakes to the population learning dynamics leads to selection much like in the works cited above. In this section I will assume that $\lambda_1, \eta > 0$.

The key idea is that mistakes must remain transient; it must not be possible for a small number of mistakes to propagate through the system and to induce large effects. This property, *transience of mistakes,* is shared by the dynamics of Kandori, Mailath and Rob, Nöldeke and Samuelson, etc..

The main result of this section relies on a property of Markov chains with stationary transition probabilities which was established by Young [1993]; Freidlin and Wentzell [1984] establish a similar property for a different class of dynamics.

Consider a Markov chain on a finite state space Ω with stationary transition probability ϕ_0. Assume that with high probability the process follows ϕ_0, but with some probability agents make mistakes. Let the corresponding noisy transition probability be denoted by ϕ_η where η is a parameter measuring the overall level of noise in the system.

Assume that ϕ_η satisfies the following three properties:

1. ϕ_η is aperiodic and irreducible for all $\eta \in (0, \bar{\eta}]$,

2. $\lim_{\eta \to 0} \phi_\eta(\tau|\sigma) = \phi_0(\tau|\sigma),\ \forall \sigma, \tau \in \Omega$, and

3. $\phi_\eta > 0$ for some η implies $\exists r \geq 0 : 0 < \lim_{\eta \to 0} \eta^{-r} \phi_\eta(\tau|\sigma) < \infty$.

It is well known that the first property implies that ϕ_η has a unique stationary distribution, and that this stationary distribution describes the long-run behavior of the dynamic irrespective of initial conditions. For any stationary distribution μ^0 of ϕ_0 let μ^0_σ denote the probability assigned to the state σ by μ^0.

If the transition from σ to τ is not impossible under ϕ_η, $r(\tau, \sigma) = r$ is called the *resistance* of the transition from σ to τ. Let $\Omega_1, \Omega_2, ..., \Omega_J$ denote the recurrent communication classes of ϕ_0. For all $i, j,\ i \neq j$, let $r_{i,j}$ be the least resistance among all directed paths beginning in Ω_i and ending in Ω_j. Define a graph $\mathcal{G}$ with vertices indexed by $\{1, 2, ..., J\}$ and for each i, j-pair a directed edge (i, j) with weight r_{ij}. A j-tree in $\mathcal{G}$ is a spanning subtree of $\mathcal{G}$, i.e., for every vertex $i \neq j$ there exists exactly one directed path from i to j. The total resistance of a j-tree is the sum of the resistances of the directed edges in that tree. The least total resistance among all j-trees, denoted γ_j, is the stochastic potential of the recurrent communication class Ω_j. Young proves the following proposition.

Let μ^η be the unique stationary distribution of ϕ_η, for any η. Then,

1. *as $\eta \to 0$, μ^η converges to a stationary distribution μ^0 of ϕ_0, and*

2. *σ is stochastically stable ($\mu^0_\sigma > 0$) if and only if σ is an element of the recurrent communication class with minimum stochastic potential.*

Note that the learning process we examine in this paper is aperiodic and irreducible, as long as players have no strategies which are never a best reply against any beliefs; even if there are players with such strategies, the result continues to hold, because the perturbed process has only one recurrent communication class. Young's proposition and Corollary 2 imply:

Proposition 4 *If $\lambda_1, \lambda_2, \lambda_3, \eta > 0$, then, $\exists \bar{\epsilon} > 0$, such that $\forall 0 < \epsilon < \bar{\epsilon}$, $\exists \bar{N}$ such that for $N_{\min} > \bar{N}$, σ is stochastically stable if and only if it belongs to a minimal-curb-state set with minimal stochastic potential.*

Also, a moment's reflection shows that for any two states σ and τ, the resistance $r(\tau, \sigma)$, if it is finite, is equal to the minimum number of mistakes needed to move from σ to τ. Recall that for the dynamic proposed here the role of mistakes is to activate certain best replies of players. The mistakes themselves do not move the system. If there are sufficiently many mistakes, a strategy may become a best reply that wasn't before; more mistakes, of the right kind, will achieve the same effect.

I will demonstrate selection among minimal curb sets for a class of two-player games. This is the class of games with two minimal curb sets, Q^1 and Q^2, such that each strategy of each player is in the projection of at least one minimal curb set. In that case I will say that the two minimal curb sets are *exhaustive.*

From Corollary 2 and Proposition 4, we know that if λ_1, λ_2, and λ_3 are all positive, then there are two recurrent communication classes of the unperturbed process, Ω_1 and Ω_2, corresponding to the two minimal curb sets, and the limit stationary distribution will assign positive weight only to the states in the minimal-curb-state set with minimum stochastic potential.

In order to find out which Ω_j is selected, we need to calculate the paths of least resistance from Ω_1 to Ω_2 and vice versa. We can move from Ω_1 to Ω_2 whenever sufficiently many type-1 (or type-2) players make a mistake and use a strategy in Q^2. Sufficiently many mistakes of the right kind eventually turn actions in Q^2 into best replies. Once the current state is supported on strategies from both curb sets, the unperturbed component of the process takes over and moves the state into either one the of the two minimal-curb-state sets with positive probability. This property, that any state not belonging to Ω_1 or

Ω_2 belongs to the basins of attraction of both minimal-curb-state sets, is a consequence of sampling (not necessarily with replacement). The least number of mistakes needed to transit from Ω_1 to Ω_2 can be expressed in terms of the players' beliefs. Let β_p be player p's belief and $\beta_p(Q^j_{-p})$ the probability which player p's beliefs assign to Q^j_{-p}. Define

$$\beta^j_p := \min\{\beta_p(Q^j_{-p}) | BR(\beta_p) \cap Q^j_p \neq \emptyset\}.$$

This is the least probability player p can attach to the set of strategies Q^j_{-p} and still have a best reply in Q^j_p. For simplicity assume that population sizes are the same and equal to N, and for any real number x let $[x]$ be the smallest integer greater than or equal to x. The least number of mistakes needed to transit from Ω_i to Ω_j is then equal to

$$\min\{[\beta^j_1 N], [\beta^j_2 N]\}.$$

This observation uses the fact that any state σ with positive support on strategies from both curb sets belongs to the basins of attraction of both curb sets. Define

$$\beta^j := \min\{\beta^j_1, \beta^j_2\}.$$

With these preliminaries we have the following result:

Proposition 5 *If G is a two-player game, with two exhaustive minimal curb sets Q^1 and Q^2, if $\lambda_1, \lambda_2, \lambda_3, \eta > 0$, then, $\exists \bar{\epsilon} > 0$, such that $\forall 0 < \epsilon < \bar{\epsilon}$, $\exists \bar{N}$ such that for $N, N_{\min} > \bar{N}$, σ is stochastically stable if and only if $\sigma \in \Omega_j$, and $\beta^j = \min\{\beta^1, \beta^2\}$.*

In the case where G is a symmetric game and the two curb sets are strict equilibria, the selection criterion in the theorem reduces to the familiar risk dominance criterion of Harsanyi and Selten [1988].

5 Examples

Consider the following example

5,1	1,5	0,0
1,5	5,1	0,0
0,0	0,0	2,2

In this example the upper-left hand curb set will be selected by the population learning dynamic. After all both players can guarantee themselves a payoff of 3 against any beliefs concentrated on this set which shows that more mistakes are needed to upset a state supported on this curb set than a state where all agents use their third strategy. In this example, a specification of the underlying deterministic process in Kandori, Mailath and Rob's (KMR) [1993] dynamic such that adjustment speeds are equal in both dimensions in regions where the basins of attraction of the equilibria overlap would yield the same selection. This shows once more that sampling with replacement is not a necessary condition for selection among curb sets which are not strict equilibria.

Note that the selection we obtain here does not depend on the values of λ_1, λ_2 and λ_3, as long as they are all positive. Furthermore, these values can be different across populations, or even within a population. Thus there is a wide range of mistake-free dynamics which yield the same selection. This is a result of the fact that the basins of attraction of different curb sets overlap. In the above example and in 2×2-coordination games it is the case that from any state not supported entirely on one of the two curb sets either of the two curb sets can be reached without mistakes. This phenomenon is already noted in KMR's paper. They also point out that there are dynamics in the two-population scenario which satisfy their *Darwinian* condition, that only the best strategy in a population grows, yet select the $(2, 2)$-equilibrium. This would be the case, for example, if in a region where the basins of attraction overlap the speed of adjustment toward one equilibrium is much faster than toward the other. Thus in the framework presented here selection of and among curb sets is obtained obtained under a large set of conditions. Outside of this framework it does matter how one formulates the mistake-free process and how exactly one embeds the mistake process into it.

That a dynamic satisfy the Darwinian condition is no guarantee that curb sets will be selected, as the next example shows.[5]

x,x	2,2	2,2	2,2
2,2	5,0	0,5	0,0
2,2	0,0	5,0	0,5
2,2	0,5	0,0	5,0

Let $x > 2$ and close to 2. The game has a unique equilibrium, with payoff vector (x, x). The unique minimal curb set coincides with this equilibrium.

Consider a dynamic in which agents from two populations are randomly matched to play this game. As they play, they make mistakes with probability η. When they make a mistake, they put strictly positive probability on each of their strategies. After each round of play they learn the true distribution of play in the last period and move to a best reply against this distribution. In the framework of this paper, this corresponds to the case of $\lambda_1 = 1$. It is easily checked that for $\eta = 0$ this Markov process has three recurrent communication classes, one corresponding to the equilibrium, one to a cycle in which agents use only their last three strategies, and one to a cycle in which the payoff vector is always $(2, 2)$.

It is also easily checked that with x close to 2, the limiting stat