Fast convergence in evolutionary models: A Lyapunov approach

Glenn Ellison, Drew Fudenberg, Lorens A. Imhof

January 26, 2015

Abstract

Evolutionary models in which \( N \) players are repeatedly matched to play a game have “fast convergence” to a set \( A \) if the models both reach \( A \) quickly and leave \( A \) slowly, where “quickly” and “slowly” refer to whether the expected hitting and exit times remain bounded when \( N \) tends to infinity. We provide simple and general Lyapunov criteria which are sufficient for reaching quickly and leaving slowly. We use these criteria to determine aspects of learning models that promote fast convergence.

Keywords: hitting time, learning model, Lyapunov function, Markov chain, recency

1 Introduction

In this paper we provide a Lyapunov-function condition for a finite-population evolutionary process to reach various target sets “quickly” in the sense that the expected waiting time is uniformly bounded over all initial conditions and population sizes. We also provide a complementary condition under which the process leaves the target set “slowly,” in the sense that the probability of getting more than \( \epsilon \) away from the set in any fixed time goes to zero as the population size increases. As this latter property seems necessary for convergence to the target set to be interesting, we only say that there is fast convergence when both conditions hold.

Our starting point is a collection of time-homogeneous Markov chains \( S^N = \{ S^N(t) : t = 0, \frac{1}{N}, \frac{2}{N}, \ldots \} \) with finite state space \( \Omega_N \) that can vary with the parameter \( N \), which we will use to index the number of players in the population. These Markov chains may track

\*We thank Ryota Iijima, Gabriel Kreindler, Larry Samuelson, and Bill Sandholm for helpful comments and National Science Foundation grant SES-1258665 for financial support.

†Massachusetts Institute of Technology, Harvard University, and Bonn University. E-mail: gellison@mit.edu, dfudenberg@harvard.edu, limhof@uni-bonn.de.
for example the play of each player at each location of a network. We then define functions \( \phi_N \) on \( \Omega_N \) that map to a space \( \Delta \) of fixed dimension, and consider the processes given by \( X^N(t) = \phi_N(S^N(t)) \). For example, these processes can describe the frequency of agents using each of a finite number of pure strategies in a game.

We present a pair of general results that use a Lyapunov function \( V \) to provide sufficient conditions for “reaching quickly” and “leaving slowly.” The first proposition says roughly that the system reaches a subset \( A \) of \( \Delta \) quickly if \( V(X^N(t)) \) decreases in expectation at rate at least \( c > 0 \) whenever the state \( X^N(t) \) is outside \( A \). Intuitively, if \( V(X^N(t)) \) is declining at rate \( c \) in expectation, then if it is currently at \( x \) it should reach \( A \) in about \( V(x)/c \) periods. The second proposition provides a closely related condition for leaving \( A \) slowly: If the maximum rate at which the state can move in a single step is bounded and \( V(X^N(t)) \) is decreasing in expectation at rate \( c \) whenever \( X^N(t) \) is outside \( A \), then the system will leave \( A \) slowly. Intuitively, this is because getting more than \( \epsilon \) away from \( A \) would take large number of steps against the drift. We also consider extensions to the case where the expected decrease of the Lyapunov function decreases as the limit is approached, and to the case where a set is reached quickly through a multi-step process that need not monotonically approach the limit.

Section 4 uses our general results to examine whether convergence is fast in various evolutionary models. The first subsection examines models in which the evolution of the process depends only on the fractions of players using each pure strategy, as in Kreindler and Young’s (2013) study of evolution with logistic choice. Here we show that there will be fast convergence to a neighborhood of the state in which all agents use one particular strategy whenever the popularity of that strategy is increasing in expectation when the state is outside the neighborhood. Such an argument is sufficient to establish fast convergence under a variety of different specifications of how strategies are revised, including stochastic choice, observation errors, and beliefs generated by random sampling. In some cases there is fast convergence if choice is sufficiently noisy (relative to the degree of risk dominance) while in others there is fast convergence for all noise levels. The second subsection discusses models that are naturally analyzed using more complex state spaces. Here it is particularly advantageous that our Lyapunov results let us work directly with random models (as opposed to taking a two-step approach of establishing that the model converges in a sufficient sense to some deterministic limit dynamic as \( N \to \infty \), and showing that the limit dynamic reaches a limit quickly.) One application here is a finite-memory fictitious play model in which we show that there is fast convergence to a neighborhood of the risk dominant equilibrium if players place enough weight on their most recent observations. The second is a
model with both local and uniform interactions in which there is fast convergence to the set where most players use the risk dominant action if the noise level is above a specified threshold.

2 Model and definitions

Suppose that for each integer \( N = 1, 2, \ldots \) we are given a discrete time homogeneous Markov chain \( S^N = \{ S^N(t) : t = 0, \frac{1}{N}, \frac{2}{N}, \ldots \} \) with finite state space \( \Omega_N. \) Let \( \Delta \) be the \((m - 1)\)-dimensional unit simplex for some fixed integer \( m > 1, \) that is, \( \Delta = \{ x \in [0, 1]^m : x_1 + \cdots + x_m = 1 \} \). In our applications to models of populations playing a game with \( m \) pure strategies, a point \( x \in \Delta \) gives the fraction of players using each of the \( m \) strategies. Given some function \( \phi_N : \Omega_N \to \Delta, \) let \( X^N \) be the stochastic process on \( \Delta \) determined by \( X^N(t) = \phi_N(S^N(t)). \) We denote the conditional probability and the conditional expectation given \( S^N(0) = s \) by \( P_s \) and \( E_s. \) For every \( A \subset \Delta \) let \( \bar{A} \) denote its closure.

To discuss the speed with which the process \( X^N \) reaches the set \( A \subset \Delta \) we write

\[
\tau_A^N = \inf\{ t \geq 0 : X^N(t) \in A \}
\]

for the random variable giving the time at which \( A \) is first reached and define

\[
W^N(A, s) = E_s(\tau_A^N)
\]

to be the expected wait to reach \( A \) conditional on the process \( S^N \) starting at state \( s \). Our first main definition is

**Definition 1.** The family \( \{ X^N \} \) reaches \( A \) quickly if

\[
\lim \sup_{N \to \infty} \sup_{s \in \Omega_N} W^N(A, s) < \infty.
\]

Note that as in Kreindler and Young (2013) this is an asymptotic property meaning that the expected waiting time remains bounded (uniformly over all starting points) in the \( N \to \infty \) limit.

Our second main definition is

**Definition 2.** The family \( \{ X^N \} \) leaves \( A \) slowly if for any finite \( T \) and for every open set \( U \) containing \( \bar{A} \)

\[
\lim_{N \to \infty} \max_{s \in \phi_N^{-1}(A)} P_s(\tau_A^N \leq T) = 0.
\] (1)

\(^1\)We believe that the assumption that \( S^N \) has a finite state space is not critical; we use it as a convenient way to avoid technical difficulties.
Note that the requirement in the definition is stronger than requiring that

\[
\lim_{N \to \infty} \min_{s \in \phi_N^{-1}(A)} W^N(\Delta \setminus U, s) = \infty.
\]

We made the definition more demanding in this dimension so that we will not count a model as leaving a set slowly just because there is some probability of being trapped within \(A\) for a very long time. Instead, it must be the case that even for very large \(T\), the probability of escaping within \(T\) periods vanishes in the \(N \to \infty\) limit.\(^2\)

Finally, we define “fast convergence” as the combination of these two properties.

**Definition 3.** The family \(\{X^N\}\) has fast convergence to \(A\) if \(\{X^N\}\) reaches \(A\) quickly and leaves \(A\) slowly.

### 3 Lyapunov criteria

This section contains several sufficient conditions for “reaching quickly” and “leaving slowly.”

#### 3.1 Main results

We first present results relating the two components of fast convergence to the existence of Lyapunov functions satisfying certain properties. Proposition 1 contains a Lyapunov condition for \(\{X^N\}\) to reach a given set quickly. Where applicable, it also provides an explicit upper bound for the expected hitting time.

To provide some intuition for the result, suppose that the Markov processes are deterministic and there is a nonnegative function \(V\) for which \(V(X^N(t))\) decreases by at least \(c/N\) in the next \(1/N\)-length time interval whenever \(X^N(t)\) is outside \(A\). Clearly, when such a process starts at \(x\) it must reach \(A\) within \(V(x)/c\) units of time. The proposition extends this simple argument to the case when the Markov process is not deterministic, but \(V(X^N(t))\) still decreases in expectation at rate \(c\).

**Proposition 1.** Let \(A \subset \Delta\), \(c \in (0, \infty)\), and let \(V : \Delta \to [0, \infty)\) be a bounded function. If

\[
E_s [V(X^N(1/N)) - V(X^N(0))] \leq -\frac{c}{N} \text{ for all } s \in \phi_N^{-1}(\Delta \setminus A),
\]

\(^2\)It would have been more accurate to say \(\{X^N\}\) “is likely to remain close to \(A\) for a long time” rather than “leaves \(A\) slowly,” but we prefer the simpler wording. For example, for \(m = 2\) a deterministic system which immediately transitions from \((0, 1)\) to \((\frac{1}{N}, 1 - \frac{1}{N})\), but then immediately returns to \((0, 1)\) is defined to leave \(\{(0, 1)\}\) slowly.
then \( W^N(A, s) \leq V(\phi_N(s))/c \) for every \( s \in \Omega_N \). In particular, if (2) holds for all \( N \) sufficiently large, then \( \{X^N\} \) reaches \( A \) quickly.

**Proof.** Fix \( N \geq 1 \), define \( Y(t) = X^N(t \wedge T^N_A) \), and let \( Z(t) \) be 0 or 1 according as \( Y(t) \in A \) or \( Y(t) \not\in A \). By (2), for \( t = 0, \frac{1}{N}, \frac{2}{N}, \ldots \), we have

\[
E \left[ V(Y(t + 1/N)) - V(Y(t)) \mid Y(t) = y, S^N(t) = s_t \right] \leq -\frac{c}{N},
\]

for any \((y, s_t)\) with \( y \not\in A \) that occurs with positive probability. The LHS of the above inequality is identically zero if \( y \in A \). So we can combine these two observations and write

\[
E \left[ V(Y(t + 1/N)) - V(Y(t)) \mid Y(t), S^N(t) \right] \leq -\frac{c}{N} Z(t).
\]

The expected change in the value function conditional on the state \( s \) at \( t = 0 \) can be computed by iterated expectations as

\[
E_s [V(Y(t + 1/N)) - V(Y(t))] = E_s \left[ E \left[ V(Y(t + 1/N)) - V(Y(t)) \mid Y(t), S^N(t) \right] \right].
\]

This gives

\[
E_s [V(Y(t + 1/N)) - V(Y(t))] \leq -\frac{c}{N} E_s [Z(t)],
\]

which is equivalent to

\[
\frac{1}{N} E_s [Z(t)] \leq \frac{1}{c} \left( E_s [V(Y(t))] - E_s [V(Y(t + 1/N))] \right).
\]

The expected time to reach \( A \) is given by \( \frac{1}{N} E_s [\sum_{k=0}^{\infty} Z(k/N)] \). The partial sums are bounded above:

\[
\frac{1}{N} \sum_{k=0}^{m-1} E_s Z \left( \frac{k}{N} \right) \leq \frac{1}{c} \sum_{k=0}^{m-1} \left( E_s [V(Y(k/N))] - E_s [V(Y((k+1)/N))] \right)
\]

\[
= \frac{1}{c} \left( E_s [V(Y(0))] - E_s [V(Y(m/N))] \right)
\]

\[
\leq \frac{1}{c} E_s [V(Y(0))] = \frac{1}{c} V(\phi_N(s)).
\]

Hence, by monotone convergence, \( W^N(A, s) \leq V(\phi_N(s))/c. \)

**Remark.** The proof resembles that of Foster’s theorem (e.g. Brémaud (1999)). Like that proof, the one here has the flavor of martingale arguments although it does not appeal to martingale results.

Our second proposition provides a criterion for \( \{X^N\} \) to leave a set \( A \) slowly. It requires that the Lyapunov function decrease in expectation whenever \( X^N(t) \) is slightly outside \( A \).
This suffices because we also assume that there is an upper bound on the rate at which the process can move. As a result, whenever the process does jump out of \( A \) it first reaches a point slightly outside \( A \). The Lyapunov condition then ensures that for large \( N \) the system is unlikely to escape the neighborhood before being drawn back into \( A \).

**Proposition 2.** Suppose there is a constant \( K < \infty \) such that \( P_s(\|X^N(\frac{1}{N}) - X^N(0)\| \leq \frac{K}{N}) = 1 \) for all \( N \) and all \( s \in \Omega_N \), where \( \| \cdot \| \) is the Euclidean norm on \( \Delta \). Let \( A \subset \Delta \) be a set with \( \phi_N^{-1}(A) \neq \emptyset \) for all \( N \geq N_0 \) for some \( N_0 \). Let \( c \in (0, \infty) \) and let \( V : \Delta \to \mathbb{R} \) be a Lipschitz continuous function such that

\[
V(x) < V(y) \quad \text{for all } x \in \bar{A}, y \in \Delta \setminus \bar{A}. \tag{3}
\]

Suppose there is an open set \( U_0 \subset \Delta \) that contains \( \bar{A} \) and

\[
E_s \left[ V \left( X^N \left( \frac{1}{N} \right) \right) - V \left( X^N(0) \right) \right] \leq -\frac{c}{N} \tag{4}
\]

whenever \( s \in \phi_N^{-1}(U_0 \setminus \bar{A}) \) and \( N \geq N_0 \). Then \( \{X^N\} \) leaves \( A \) slowly.

**Remark.** In the Appendix we prove the stronger result that there is a constant \( \gamma \) such that for \( N \) large the exit time \( \tau_{\Delta \setminus U}^N \) first order stochastically dominates an exponential random variable with mean \( e^{\gamma N} \). The argument writes the exit time as the sum over all instances that the process reaches \( \bar{A} \) of the transit times to either reach \( \Delta \setminus U \) or to return to \( \bar{A} \). Each such transit takes at least one unit of time, so it suffices to show that the number of transits is large enough. The key step is to show that the probability of reaching \( \Delta \setminus U \) before returning to \( \bar{A} \) when starting in \( \bar{A} \) is small (and declines exponentially in \( N \)). Intuitively, this is true because the bound on the speed with which the process can move implies that any initial step out of \( \bar{A} \) reaches a point that is bounded away from \( \Delta \setminus U \), and because the drift is toward \( \bar{A} \) and many steps against the drift are needed to reach \( \Delta \setminus U \), it becomes increasingly unlikely that a transit will lead to \( \Delta \setminus U \) rather than to \( \bar{A} \) when \( N \) is large.

The technical argument focuses on the random variable \( Y^N(t) = e^{\delta_0 N V(X^N(t))} \). Using the hypotheses that \( X^N(t) \) moves in bounded steps and \( V \) is Lipschitz continuous, we can approximate the exponential in the definition of \( Y^N \) using a Taylor expansion and show that for an appropriately small value of \( \delta_0 \), the random variable \( Y^N(t) \) is a positive supermartingale whenever \( X^N(t) \) is between \( \bar{A} \) and \( \Delta \setminus U \). This allows us to apply general results on positive supermartingales to conclude that the probability of reaching \( \Delta \setminus U \) before returning to \( \bar{A} \) is at most \( e^{-\gamma N} \) for some constant \( \gamma \).

\(^3\)For example, this would hold if at most \( K \) players change their strategies in each \( 1/N \)-length time interval.
The hypotheses of Propositions 1 and 2 are similar. One difference is that Proposition 2 is more demanding in that it has added a bound on the speed with which the process can move. Another is that Proposition 1 requires that the Lyapunov condition holds on a larger set (whenever $X^N(t)$ is outside $A$ versus just when $X^N(t)$ is in some neighborhood of $\bar{A}$). When the more restrictive version of each hypothesis holds the process will both reach $A$ quickly and leave $A$ slowly. Hence, we have fast convergence to $A$.

**Corollary 2.1.** Suppose the hypotheses of Proposition 2 are satisfied and that condition (4) holds also for all $s \in \phi_N^{-1}(\Delta \setminus A)$ when $N \geq N_0$. Then $\{X^N\}$ has fast convergence to $A$.

**Proof.** Any Lipschitz continuous function on $\Delta$ is bounded, and replacing $V(x)$ by $V(x) - \min_{y \in \Delta} V(y)$ if necessary, one may assume that $V$ is nonnegative. Hence, the hypotheses of Propositions 1 and 2 are satisfied. □

### 3.2 Extensions

Some models will not satisfy the hypotheses of the results above because the expected decrease in the Lyapunov function decreases to 0 as the state approaches the target set $A$. In the case of Proposition 1, one will often be able to slightly weaken the desired conclusion and argue that for any open neighborhood $U$ of $A$, the model reaches $U$ quickly. This will follow if there is a positive lower bound on the rate at which the Lyapunov function decreases when $x \in \Delta \setminus U$. In the case of Proposition 2 we can do even better because a drift that vanishes at $A$ is not a problem. To describe this formally, for $A \subset \Delta$, let $U(A, \epsilon)$ denote the $\epsilon$-neighborhood of $A$ in $\Delta$, $U(A, \epsilon) = \{x \in \Delta : \inf_{y \in A} \|x - y\| < \epsilon\}$.

**Proposition 3.** The conclusion of Proposition 2 that $\{X^N\}$ leaves $A$ slowly remains true if the Lyapunov hypothesis is replaced with “Suppose there is an open set $U_0 \subset \Delta$ that contains $\bar{A}$ and for every $\epsilon > 0$ there are numbers $c_\epsilon > 0$ and $N_\epsilon \geq N_0$ such that

$$E_s \left[ V \left( X^N \left( \frac{1}{N} \right) \right) - V \left( X^{(N)}(0) \right) \right] \leq -c_\epsilon \frac{N}{N}$$

for all $s \in \phi_N^{-1}(U_0 \setminus U(\bar{A}, \epsilon))$ and $N \geq N_\epsilon$.”

**Proof.** Let $U$ be an open set containing $\bar{A}$. Let $v_1 := \max_{x \in \bar{A}} V(x)$ and choose $v_2 > v_1$ so small that $\tilde{A} := \{x \in \Delta : V(x) \leq v_2\} \subset U \cap U_0$. Then choose $\epsilon > 0$ so that $U(\bar{A}, \epsilon) \subset \tilde{A}$. Now apply Proposition 2 with $A, c, N_0$ replaced by $\tilde{A}$, $c_\epsilon$ and $N_\epsilon$. This yields that $\{X^N\}$ leaves $\tilde{A}$ slowly. Thus, as $A \subset \tilde{A} \subset U$,

$$\lim_{N \to \infty} \max_{s \in \phi_N^{-1}(A)} P_s(\tau^N_{\Delta \setminus U} \leq T) \leq \lim_{N \to \infty} \max_{s \in \phi_N^{-1}(\tilde{A})} P_s(\tau^N_{\Delta \setminus U} \leq T) = 0$$
for all $T \in [0, \infty)$. This proves that $\{X^N\}$ leaves $A$ slowly. □

In some cases models have fast convergence even though they do not always “drift” toward the selected set. For example, in a model where $N$ agents are arranged on a circle to play a $2 \times 2$ coordination game and play a best response to the average play of their $2k$ closest neighbors unless a mutation occurs, the number of players playing the risk-dominant action is not monotonically increasing – it can decrease in expectation if there is not a sufficiently large cluster of players playing the risk-dominant action. Ellison (1993) nonetheless shows that play reaches a neighborhood of the state where everyone plays the risk-dominant equilibrium quickly. Intuitively, this occurs because evolution can proceed in a two-step manner: Each period there is a nonzero chance that a cluster of players using the risk-dominant action will form, and whenever such a cluster exists, the model drifts toward everyone playing the risk-dominant equilibrium.

Our results can be extended so that they apply to some models of this variety. Specifically, the proposition below shows that $\{X^N\}$ reaches a set $A$ quickly when three conditions hold: (1) $\{X^N\}$ reaches a superset $B$ of $A$ quickly; (2) $\{X^N\}$ does not stay too long in $B \setminus A$, that is, $\{X^N\}$ reaches $A \cup B^c$ quickly; and (3) the probability that $\{X^N\}$ reaches $A$ before $B^c$ when starting anywhere in $B \setminus A$ is bounded away from zero. The proof uses an argument related to Wald’s equation for the expectation of a random sum of i.i.d. random variables. A transition from $B^c$ to $A$ consists of a random number of transitions from $B^c$ to $B \setminus A$ and back to $B^c$ with a final transition to $A$. The assumptions ensure that the expected lengths of the individual transitions are bounded and that the number of these transitions has a finite expectation as well.

**Proposition 4.** Let $A \subset B \subset \Delta$. Suppose $\{X^N\}$ reaches $B$ quickly, $\{X^N\}$ reaches $A \cup B^c$ quickly, and there exist $c > 0$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$,

$$P_s(X^N(\tau_{A \cup B^c}^N) \in A) \geq c \quad \text{for all } s \in \phi^{-1}_N(B \setminus A).$$

(5)

Then $\{X^N\}$ reaches $A$ quickly.

The proof is contained in the Appendix. We will use this result in the proof of Proposition 6 in our applications section.

4 Applications

This section analyzes two sets of applications to non-equilibrium game dynamics. Subsection 4.1 analyzes models where the evolution of the system depends only on the number
of agents using each strategy, so that the state space \( \Omega_N \) can simply be these numbers. Here, we connect fast convergence with whether the popularity of some strategy is increasing in expectation, and derive several results about particular models as corollaries. Subsection 4.2 considers two examples in which the natural state spaces are more complex: one is a model of personal experience learning; the other is a model with local and global interactions.

### 4.1 Dynamics based on population averages and random samples

Suppose that a population of \( N \) players are choosing strategies for a game with \( m \) pure strategies at \( t = 0, \frac{1}{N}, \frac{2}{N}, \ldots \). Throughout this section we assume that in each time period a single agent is chosen at random to revise his strategy. (With a time renormalization this could describe a model in which revision opportunities arrive in continuous time according to independent Poisson processes.) Let \( X_i^N(t) \) denote the fraction of agents that play strategy \( i \) at time \( t \). If the current state of the population is \( x \), suppose the revising agent chooses strategy \( i \) with probability \( f_i(x) \), regardless of the agent’s own current action.\(^4\) For now, we let \( f_1, \ldots, f_m : \Delta \to [0,1] \) be arbitrary functions with \( f_1(x) + \cdots + f_m(x) = 1 \) for all \( x \).

Part (a) of our general result shows that if the probability of choosing some strategy \( i \) exceeds its current share whenever that share is below some threshold \( a \), then there is fast convergence to the states where the share of \( i \) exceeds \( a \). Part (b) gives a partial converse.

**Proposition 5.** Consider a dynamic \( \{X^N\} \) with choice rules \( f_1(x), \ldots, f_m(x) \). Let \( a, c \in (0, 1) \) and \( i \in \{1, \ldots, m\} \).

(a) If

\[
  f_i(x) - x_i \geq c \text{ whenever } x_i \leq a,
\]

then \( \{X^N\} \) has fast convergence to \( \{x : x_i > a\} \) and \( W^N(\{x : x_i > a\}, X^N(0)) \leq \frac{1}{c} \).

(b) If \( f_i \) is continuous and

\[
  f_i(x) < a \text{ whenever } x_i = a,
\]

then \( \{X^N\} \) leaves \( \{x : x_i < a\} \) slowly.

\(^4\)The assumption that the \( f_i \) do not depend on the current action rules out models in which agents respond to the current play of all others not including themselves, but since the effect of any one agent’s strategy on the overall state is of order \( 1/N \), we do not expect it to matter for large \( N \) except perhaps in knife-edge cases.
Proof. Here $\phi_N(x) = x$, $S_N = X_N$, and $\Omega_N = \{x \in \Delta : N x \in \mathbb{N}^m\}$.

(a) Let $V(x) = 1 - x_i$. Fast convergence to $\{x : x_i > a\}$ then follows from Corollary 2.1 since for every $x \in \Omega_N$ with $x_i \leq a$,

$$E_x \left[ V \left( X_N \left( \frac{1}{N} \right) \right) - V \left( X_N(0) \right) \right] = -E_x \left[ X_i^N \left( \frac{1}{N} \right) - X_i^N(0) \right] = -\frac{1}{N}(1 - x_i)f_i(x) + \frac{1}{N}x_i[1 - f_i(x)] = -\frac{1}{N}[f_i(x) - x_i] \leq -\frac{c}{N}.$$

The uniform bound on the convergence time follows from Proposition 1.

(b) Set $c' = -\frac{1}{2}\sup\{f_i(x) - x_i : x \in \Delta, x_i = a\}$. The set of $x$ with $x_i = a$ is compact so $c' > 0$. The uniform continuity of $f_i$ implies that we can choose an $\epsilon > 0$ so that $f_i(x) - x_i \leq -c'$ whenever $x_i \in [a, a + \epsilon]$. The assertion follows from Proposition 2 with $V(x) = x_i$. □

Remark. The continuity of $f_i$ is used in part (b) only to show that $f_i(x) - x_i$ is negative and bounded away from 0 when $x_i$ is in an interval around $a$. Accordingly, part (b) could be strengthened to apply to discontinuous $f_i$ in a variety of ways. For example, it would suffice to assume that there exists a $c > 0$ and an $a' > a$ such that $f_i(x) - x_i \leq -c$ whenever $x_i \in (a, a')$. In view of Proposition 3 it is also sufficient to assume that $f_i$ is upper semicontinuous with $f_i(x) - x_i < 0$ whenever $x_i \in (a, a')$.

Now we present several examples that use this result in coordination games under different dynamics. In each case we assume that strategy 1 is risk dominant.

Example 1 generalizes Kreindler and Young (2013)'s analysis of logit responses in $2 \times 2$ games to games with $m$ strategies. Specifically, write $\pi_j : \Delta \rightarrow \mathbb{R}$ for the expected payoff function of each strategy $j$ given the state $x \in \Delta$ and suppose that for a constant $\beta \in (0, \infty)$, a revising agent chooses strategy $i$ with probability

$$f_i(x) = f_i(x, \beta) = \frac{e^{\beta \pi_i(x)}}{\sum_{j=1}^{m} e^{\beta \pi_j(x)}}.$$

Kreindler and Young (2013) show that when $\beta$ is small enough (which is a form of “enough noise” condition) in a two-strategy game the expected popularity of the risk-dominant equilibrium is monotonically increasing until a neighborhood of the state where everyone plays it is reached, and that the state reaches this neighborhood quickly. We generalize this

\footnote{Such choice rules can arise either from payoff shocks with an extreme-value distribution or from maximizing a weighted sum of expected utility and entropy. Fudenberg and Kreps (1993), Fudenberg and Levine (1995) respectively introduced these two types of general stochastic choice rules in the study of learning and fictitious play; Blume (1993) introduced the logit case in evolutionary models and noted the role of the noise level there.}
result to $m$-action coordination games and strengthen the characterization from “reaches quickly” to “fast convergence.”

Example 1. Consider the model above. Suppose that for all $x \in \Delta$ with $x_1 \leq \frac{1}{2}$,

$$\pi_1(x) \geq \alpha_1 x_1 \quad \text{and} \quad \pi_i(x) \leq \alpha_2 x_i \text{ for } i = 2, \ldots, m,$$

where

$$\alpha_1 > \frac{2(m - 2 + e^{\beta \alpha_2})}{e \beta} \quad \text{and} \quad \alpha_2 \geq 0.$$

Then the logit dynamic $\{X^N\}$ has fast convergence to $\{x \in \Delta : x_1 > \frac{1}{2}\}$.

The proof is in the Appendix; it uses the Schur-convexity of $\sum_{i=2}^{m} e^{\beta \alpha_2 x_i}$ to show that the hypothesis of Proposition 5 is satisfied. Note that this hypothesis involves the relationship between $\alpha_1$ and $\beta$: A greater degree of risk dominance is usually required if there is less noise, because otherwise play can get stuck at the dominated equilibrium. For example, in a two-action pure coordination game where the payoff to the equilibrium where both players use action 2 is $\alpha_2 = 1$, there is fast convergence to the set where the majority of players play action 1 when the payoff to coordinating on action 1 is at least 2 for $\beta = 1$, at least $e$ for $\beta = 2$, and at least $2e^2/3 \approx 5$ if $\beta = 3$.

The next example generalizes the previous one by supposing that the stochastic nature of the revision process arises from agents maximizing a perturbed utility function of the form $U(p) = \sum_i p_i \pi_i(x) - c(p_i)/\beta$ as in the literature on fictitious play, where $1/\beta$ is a measure of the amount of noise in the system, and $c : (0, 1] \to \mathbb{R}$ is continuously differentiable, strictly convex, and satisfies the Inada condition that $\lim_{p \to 0} c'(p) = -\infty$; the logit choice function corresponds to $c(p) = p \ln p$. Hofbauer and Sandholm (2002) show that all random utility models where agents maximize the sum of their payoff plus a random shock can also be regarded as members of this class, and that the class is strictly more general in that the

---

6Our assumptions on payoffs are slightly broader than pure coordination games.
7The inequality never holds when $\beta$ is in a neighborhood of 0, because then choice is almost independent of payoffs, and the risk-dominant equilibrium will only be played by a $1/m$ fraction of agents. In the $m = 2$ case even random choice will lead to half of the agents playing strategy 1 so it is very easy to get fast convergence; here Kreindler and Young (2013) show that even $\alpha_1 > \alpha_2$ suffices whenever $\beta \leq 2$.
8Fudenberg and Levine (1995) show how this leads to “stochastic fictitious play,” and generates Hannan-consistent choice, meaning that its long-run average payoff is at least as good as the best response to the time average of the moves of Nature and/or other players. Hofbauer and Sandholm (2002), Benaïm, Hofbauer, and Hopkins (2009), and Fudenberg and Takahashi (2011) use perturbed utility to construct Lyapunov functions for stochastic fictitious play, and study perturbed utility in an evolutionary model. Fudenberg, Iijima, and Strzalecki (2014) show that this form of stochastic choice corresponds to the behavior of an agent who is uncertain about his payoff function and has a form of variational preferences and so randomizes to guard against moves by a malevolent Nature.
cost function \( c(p) = -\ln p \) (and many others) do not correspond to any form of random utility. Still, we can show that in \( 2 \times 2 \) coordination games the system where revising agents act to maximize \( U(p) \) satisfies the monotonicity requirement of Proposition 5 (a) if there is enough noise and so the system has fast convergence to the set where more than \( 1/2 \) the agents use the risk-dominant strategy:

**Example 2.** Consider the model above, i.e. suppose action 1 is risk dominant in a \( 2 \times 2 \) coordination game and let \( f_1(x) = \arg \max_p \left[ p\pi_1(x) + (1-p)\pi_2(x) - (c(p) + c(1-p))/\beta \right] \) with \( c \) satisfying the conditions above. Then there is fast convergence to the set \( \{ x \in \Delta : x_1 > \frac{1}{2} \} \), provided \( \beta \) is sufficiently small.

The proof is in the Appendix and notes that the example is a special case of Proposition 7 below. It can also be derived using Proposition 5.

The next example considers a different behavior rule for revising agents: Instead of playing a stochastic best response to the current distribution of strategies, players use an exact best response to a noisy signal of the state. Specifically, at each time \( t = \frac{1}{N}, \frac{2}{N}, \ldots, \) one randomly chosen agent \( i \) draws, with replacement, a sample of size \( s \) from the current population. Assume that rather than correctly recognizing how each sampled agent \( j \) is playing, agent \( i \) with probability \( 1 - \epsilon \) correctly recognizes how \( j \) is playing, but with probability \( \epsilon \) instead thinks \( j \) is playing another action, with all strategies other than the true one being equally likely, and assume that observation errors are independent across all observations. The revising agent \( i \) then adopts the best response to this set of observations. Here, we consider more general \( m \times m \) games, but assume that strategy 1 is strongly risk dominant in the sense that agent \( i \) will choose strategy 1 if he observes at least one player playing strategy 1. Note that if \( \epsilon = 1 - \frac{1}{m} \), then each observation is equally likely to be each pure strategy irrespective of actual play.

In this model we find that there is fast convergence to the set of states where the popularity of strategy 1 exceeds a threshold. In contrast to the earlier examples of this section, we get fast convergence here even when the noise level is arbitrarily low.

**Example 3.** In the model described above, if strategy 1 is strongly risk dominant and the number of agents sampled is at least two, then \( \{X^N\} \) has fast convergence to \( \{ x : x_1 > 1 - (1 - \frac{1}{m})^s \} \) for every \( \epsilon \in (0, 1 - \frac{1}{m}) \).

We defer the calculations to the Appendix. The key once again is that the probability of choosing the risk-dominant strategy 1 is above its current share as long as the share is below the threshold. This follows from the fact that the above probability is bounded
below by a concave function of that share, with the bound being above the share at \( x_1 = 0 \) and equal to it at \( x_1 = 1 - (1 - \frac{1}{m})^s \).

As a final application of Proposition 5 we suppose that agents’ beliefs derive from correctly observing a random sample of play. Roughly, our results say that such noisy beliefs speed evolution in the sense that sampling enlarges the set of parameters for which evolution is fast. More precisely, we return to \( 2 \times 2 \) coordination games and we compare the evolution of the Markov chain \( X^N \) describing the dynamics where the revising agent chooses strategy 1 with probability \( g(x) \) if the current state of the population is \( x \) with the evolution of the chains \( X^N_s \) in which updating agents apply the same rule \( g \) to a random sample of size \( s \).

In the random sampling model the probability \( f(s)(x) \) that a revising agent chooses strategy 1 is

\[
f(s)(x) = \sum_{k=0}^{s} g\left(\frac{k}{s}\right) \left(\begin{array}{c} s \\ k \end{array}\right) x^k (1 - x)^{s-k}, \quad x \in [0, 1].
\]

We begin by considering when the implication

\[
\{X^N\} \text{ reaches } (\frac{1}{2}, 1] \text{ quickly } \Rightarrow \{X^N_s\} \text{ reaches } (\frac{\xi}{2}, 1] \text{ quickly}
\]

is true. Proposition 6 provides a symmetry condition on \( g \) under which implication (6) holds for every \( s \). This generalizes Kreindler and Young’s (2013) result beyond logit best responses and strengthens the results to show there is fast convergence in the model with sampling beliefs whenever there is fast convergence with full information (as opposed to when a sufficient condition for reaching quickly holds). To understand the condition we place on \( g \) and why it is a generalization, note that when sup\(_{0 < x < \frac{1}{2}}\) \( g(x) < \frac{1}{2} \), \( \{X^N\} \) cannot reach \( (\frac{1}{2}, 1] \) quickly. So, ignoring a knife-edge case, we assume there exists \( x^* \in (0, \frac{1}{2}) \) with \( g(x^*) \geq \frac{1}{2} \). In two-action games with strategy 1 being risk dominant, the logit model satisfies a stronger version of the symmetry condition: \( g(x^* + x) + g(x^* - x) = 1 \) for all \( x \in [0, x^*] \) where \( x^* \) is the mixed-strategy equilibrium. We loosen this to only require an inequality.

**Proposition 6.** In the above model of learning in a \( 2 \times 2 \) game let \( x^* \in (0, \frac{1}{2}) \). Suppose that \( g(x^* + x) + g(x^* - x) \geq 1 \) for all \( x \in [0, x^*] \) and that \( g \) is strictly increasing. Suppose \( \{X^N\} \) reaches \( [x^*, 1] \) quickly. Then, there exists \( \xi > \frac{1}{2} \) such that \( \{X^N\} \) and \( \{X^N_s\} \) have fast convergence to \( (\xi, 1) \) for every \( s \).

---

9Here, the state is the fraction of agents using strategy 1.

10In the analysis of the relation between \( \{X^N_s\} \) and \( \{X^N\} \) we take advantage of the fact that the functions \( f(s) \) are the Bernstein polynomials of \( g \) to exploit known results about properties of these polynomials.

11We also have strengthened (6) in that we relax the assumption that \( \{X^N\} \) reaches \( (\frac{1}{2}, 1] \) quickly to the assumption that \( \{X^N\} \) reaches \( [x^*, 1] \) quickly.
Our next result sharpens the message of the previous one by showing that there is a range of parameter values for which there is fast convergence with random sampling beliefs, but not when players observe the full state. Consider as above a family of decision rules $g(x, \beta)$, where $1/\beta$ is a measure of the level of noise.

**Proposition 7.** Consider a 2-action game with affine payoff functions $\pi_i$ and let $x^* \in (0, \frac{1}{2})$ be such that $\text{sign}(\pi_1(x) - \pi_2(x)) = \text{sign}(x - x^*)$. Suppose that the choice rule $g$ can be written in the form $g(x, \beta) = P[\beta(\pi_1(x) - \pi_2(x)) \geq \epsilon]$ where the support of $\epsilon$ is $\mathbb{R}$, $\epsilon$ and $-\epsilon$ have the same distribution, and $P[\epsilon = 0] < 1 - 2x^*$. Then there exist $0 < \beta^* < \beta^*_s$ such that

(i) if $\beta \in (0, \beta^*)$ then both models have fast convergence to the set $(\frac{1}{2}, 1]$.

(ii) if $\beta \in (\beta^*, \beta^*_s)$ then the system with random sampling has fast convergence to $(\frac{1}{2}, 1]$ but the system with full information does not.

Intuitively, fast convergence requires a sufficient amount of noise, and random sampling provides an additional stochastic element without breaking the monotonicity needed to appeal to Proposition 5. Note that the condition of this proposition applies to choice rules generated by the perturbed utility functions considered above.\(^\text{12}\)

### 4.2 More complex state spaces

In the above examples, the state space of the Markov process is naturally taken to be $\Delta_N$ as the dynamics only depend on the population shares using each strategy. Many other interesting models require a more complex state space. For example, the probability that a player adopts some strategy may depend on the player’s position in a network, who he has been matched with previously, or some other factor. An attractive feature of our Lyapunov approach is that it also applies to a variety of such models. In this subsection we first discuss two applications: one involving learning from personal experience with recency weights, and one in which agents have both local and uniform interactions.

First, consider a finite-memory fictitious-play style learning model in which $N$ agents are matched to play a two-action two-player game $G$ at $t = 0, \frac{1}{N}, \frac{2}{N}, \ldots$ and learn only from play in their own interactions.\(^\text{13}\) Assume that strategy 1 is $p$-dominant for some $p < \frac{1}{2}$ and that

\(^\text{12}\) Note also that the assumption that the support of $\epsilon$ is $\mathbb{R}$ implies that $g(x, \beta)$ is always strictly between 0 and 1, i.e. choice is always stochastic. This is needed for the result: when $c(p) = p^2$ as in Rosenthal (1989) and Voorneveld (2006) the choice rule is deterministic at some states and as we show in the Appendix the conclusion of the proposition is false. In this case, there is a $\beta^* > 0$ such that both models have fast convergence to $(\frac{1}{2}, 1]$ if $0 < \beta < \beta^*$ and the state 0 is an absorbing state for both models if $\beta \geq \beta^*$.

\(^\text{13}\) Ellison (1997) studied the finite-memory variant as well as the more traditional infinite-memory fictitious play model with the addition of a single rational player, and showed that the rational player could manipulate the play of a large population of opponents when one action was strongly risk-dominant. Most
agents remember the outcomes of their $k$ most recent matches.\textsuperscript{14} Within each period the players are randomly matched and play the game using strategies they selected at the end of the previous period. One player is then randomly selected to update his or her strategy. The updating player does one of two things. With probability $1 - \epsilon$ he selects a strategy which is a best response to a weighted average $w_1 a_{-i,t} + w_2 a_{-i,t-1} + \ldots + w_k a_{-i,t-(k-1)}$ of the actions $a_{-i,t}, \ldots, a_{-i,t-(k-1)}$ that his opponents have used in the most recent $k$ periods. With probability $\epsilon$ he selects a strategy uniformly at random. The selected strategy will be used until the player is next chosen to update.

Informally, a motivation for using recency weights is that agents would want to place more weight on more recent observations if they believed that their opponents’ play was changing over time. There is ample experimental evidence suggesting that beliefs are indeed more heavily influenced by recent observations both in decision problems and in games; see Cheung and Friedman (1997) for one of the first measures of recency bias in a game theory experiment, and Erev and Haruvy (2013) for a survey of evidence for recency effects in experimental decision problems. Benaïm, Hofbauer and Hopkins (2009) and Fudenberg and Levine (2014) provide theoretical analyses of recency, but neither consider the large-population limit that is our focus here.

We show that this model has fast convergence to a strongly risk dominant equilibrium if the weights place enough emphasis on recent actions. Note that this result does not require that noise levels be above some threshold, and that play converges to an arbitrarily small neighborhood of the selected equilibrium when the level of noise is sufficiently low. To state the result formally, note that the model defines a Markov process $S^N(t)$ on the state space $\Omega^N = \{1, 2, \ldots, m\}^{kN + N}$: the first $kN$ components of the state vector record what each player saw in the periods $t - \frac{1}{N}, t - \frac{2}{N}, \ldots, t - \frac{k}{N}$; and the last $N$ components record the action that each player has selected for use in period $t$. Given any state $S^N(t)$ we can define a random variable $X^N(t) \subset \Delta$ to be the fraction of players who have each pure strategy as their selected action in state $S^N(t)$.

**Example 4.** Consider the model above with $k > 1$. Suppose that strategy 1 is $p$-dominant in $G$ and the recency weights satisfy $w_1 > p$ and $w_2 + w_3 > p$. Then, for any $\epsilon > 0$ the model described above has fast convergence to \(\{x \in \Delta | x_1 > 1 - 1.2\epsilon\}\).

**Remarks.**

other studies of fictitious play assume that all agents in the same player role have the same beliefs. One exception is Fudenberg and Takahashi (2011), who allow each of $N$ agents to have different beliefs; they focus on the asymptotics for a fixed $N$ rather than the large-population limit.

\textsuperscript{14}That is, strategy 1 is the unique best response to every mixed strategy that assigns at least probability $p$ to strategy 1 (Morris, Rob, Shin (1995)).
1. The result implies there can be fast convergence even with long memories and moderate levels of \(p\)-dominance provided that players place substantial weights on their most recent experiences. For example, if players place weight proportional to \(\left(\frac{2}{3}\right)^{n-1}\) on their \(n\)th most recent observation, then \(w_1 > 1/(1 + 2/3 + 4/9 + \ldots) = \frac{1}{3}\) and \(w_2 + w_3 > \frac{2}{3} + 4/27 > \frac{1}{3}\), so there is fast convergence to a neighborhood of strategy 1 if strategy 1 is \(\frac{1}{3}\)-dominant.

2. Another form of recency weighting is to completely ignore all observations from more than \(k\) periods ago but weight all of the last \(k\) periods equally, so that \(w_n = \frac{1}{k}\) for \(n = 1, 2, \ldots, k\). In this case the result implies there is fast convergence to a neighborhood of the state where everyone plays strategy 1 if strategy 1 is \(\frac{1}{k}\)-dominant. This implies that we always have convergence to a risk-dominant equilibrium if memories are short enough (but longer than one period).

Proof of Example 4. Consider the two-period ahead dynamics of the model. By Propositions 1 and 2 it suffices to show that we can find \(c > 0\) for which

\[
\inf_{s \in \phi_N(x)} E[(X_1^N(2/N) - X_1^N(0))|X^N(0) = x, S^N(t) = s] \geq \frac{c}{N}
\]

for all \(x\) with \(x_1 \in [0, 1 - 1.2\epsilon]\).

We can evaluate the change in the popularity of strategy 1 by counting every time a player playing strategy 1 is selected to update as a loss of 1 and every adoption by an updating player as a gain of 1. The expected losses are \(-x_1\) from revisions at \(t = 0\) and at most \(-(x_1 + \frac{1}{N})\) from revisions at \(t = \frac{1}{N}\). There will be a gain without a mutation from the period \(t\) revision if the player selected to update at \(t\) is matched with a player who uses strategy 1 in period \(t\), or if he saw strategy 1 in both periods \(t - \frac{2}{N}\) and \(t - \frac{1}{N}\). At \(t = 0\) only the former is guaranteed to be possible for all \(s\) – the worst case state is that the matching was such that the sets of players who saw strategy 1 in periods \(t - \frac{1}{N}\) and \(t - \frac{2}{N}\) are disjoint – so all we can say is that the expected number of adoptions is at least \(x_1(1 - \epsilon)\). But the \(t = \frac{1}{N}\) revision will produce an adoption of the latter type if the player who saw strategy 1 in period \(t = -\frac{1}{N}\) is randomly matched with a player playing strategy 1 at \(t = 0\) and is then randomly selected to update at \(t = \frac{1}{N}\). So the expected number of non-mutation adoptions is at least \((x_1 - \frac{1}{N}) + (1 - x_1 + \frac{1}{N})x_1(x_1 - \frac{1}{N})(1 - \epsilon)\). And there are \(\epsilon/2\) adoptions in expectation due to mutations in each of the two periods.

Adding all of these changes together and ignoring all of the \(\frac{1}{N}\) terms, it suffices to show there is a \(c > 0\) for which

\[
-2x_1 + (2x_1 + x_1^2 - x_1^3)(1 - \epsilon) + \epsilon > c,
\]
for all $x_1 \in [1 - \ell \epsilon]$, where we have written $\ell$ for 1.2 to clarify the argument and suggest how it could be strengthened. Note that this is equivalent to

$$x_1^2 - x_1^3 > c + (-1 + 2x_1 + x_1^2 - x_1^3) \epsilon. \quad (8)$$

It is easy to choose a $c$ so that this will be satisfied for all $x_1 \in [0, \frac{1}{3}]$. For such $x_1$ we have $-1 + 2x_1 + x_1^2 - x_1^3 = -1 + 2x_1 + x_1(x_1(1 - x_1)) \leq 1 + \frac{2}{3} + \frac{2}{27} = \frac{29}{27}$, so any $c < \frac{7}{27} \epsilon$ suffices.

For $x_1 \in [\frac{1}{3}, 1 - \ell \epsilon]$ the LHS of (7) is concave so it suffices to show that for some fixed $c$ equation (7) holds for both $x_1 = \frac{1}{3}$ and $x_1 = 1 - \ell \epsilon$. We already noted a restriction on $c$ that makes (7) hold for $x_1 = \frac{1}{3}$ so it remains only to show that it is satisfied for $x_1 = 1 - \ell \epsilon$. To do this, we substitute $(1 - x_1)/\ell$ for $\epsilon$ into (8) and find that it is equivalent to

$$\frac{1 - x_1}{\ell} ((\ell x_1^2 - (-1 + 2x_1 + x_1^2 - x_1^3)) > c.$$  

A numerical calculation shows that the polynomial $(1 - 2x_1 + (\ell - 1)x_1^2 + x_1^3)$ is positive for all $x_1 \in [0, 1]$ if $\ell > 1.15$, in which case one can (given any fixed $\epsilon$) choose a $c$ that satisfies (8) for all $x_1 \in [\frac{1}{3}, 1 - \ell \epsilon]$. □

**Remark.** The result could be strengthened to show that there is fast convergence to a somewhat smaller set by considering $k$-period ahead transitions instead of the two-period ahead transitions considered in the proof.

Second, consider the following local interaction model with a “small-world” element: $N$ players are arranged around circle and are randomly matched to play a $2 \times 2$ coordination game. One player chosen at random considers updating his strategy in each period. With probability $1 - \epsilon > 1/2$ the updating player plays a best response to the average of the play in the previous period of four players: his two immediate neighbors and two players selected uniformly at random, with probability $\epsilon$ the updating player chooses the opposite strategy.

We focus on the case where strategy 1 is risk-dominant, but not $\frac{1}{4}$-dominant: an updating player will choose strategy 1 if at least two of the four players in his sample are using it, but will use strategy 2 if at least 3 players in the sample are using strategy 2.

An application of our Lyapunov condition shows that the model will have fast convergence to the risk dominant equilibrium if the noise level is above a critical threshold which turns out to be not very large, even though clusters of players using strategy 1 in a

---

15To simplify the algebra we assume that the random samples are independent draws from the full population (including the player himself and the immediate neighbors).
population of mostly strategy 2 players tend to shrink slowly from the edges. Intuitively, when \( x_1 \) is very small the drift is to the right because of the \( \epsilon \) noise. When \( x_1 \) is larger there are additional adoptions from players who have neighbors playing strategy 1 and/or see it in their random sample. This force dominates until almost all players are playing the risk dominant action.

**Example 5.** Consider the model above. Suppose that strategy 1 is risk dominant. Then, for \( \epsilon > 0.065 \) the model has fast convergence to \( \{ x \mid x_1 > 1 - 1.5\epsilon \} \).

The basic structure of the argument is similar to that for Example 4. The proof is in the Appendix. Another application that we do not develop here is the use of Proposition 4 to develop results about the selection of iterated \( p \)-dominant equilibria. We believe that this would lead to results like those in Oyama, Sandholm, and Tercieux (2014). In addition, we conjecture that our results could be fruitfully applied to the network models of Kreindler and Young (2014), where the dimension of the state space is the size of the population.

## 5 Related Literature

Ellison (1993) raises the issue of slow convergence; it shows that in the model of Kandori, Mailath, and Rob (1993) the rate of convergence slows as the noise level converges to zero, and that the expected waiting time to reach the long-run equilibrium grows exponentially as the population size \( N \) increases. Blume (1993) provides conditions under which an infinite-population local interaction model is ergodic; evolution can be thought of as fast for large \( N \) if it still occurs when \( N = \infty \). Möbius (2000) defines a model as “clustering” on a set of states \( A \) if the probability of \( A \) under the ergodic distribution converges to 1 as \( N \to \infty \), and analyzes the limit of the worst case expected wait to reach \( A \). Kreindler and Young (2013) defines a concept of “fast selection” which is roughly equivalent to what we call “reaches \( A \) quickly.” Relative to this the novelty in our concept consists of adding the “leaves \( A \) slowly” requirement.

There is also a sizable literature on factors that promote fast evolution, including local interaction (Ellison (1993) and Blume (1993)), endogenous location choice (Ely (2002)), homophily (Golub and Jackson (2012)), and population structure (Hauert, Chen and Imhof (2014)). Samuelson (1994) and Ellison (2000) note that evolution is more rapid when it can proceed via a series of smaller steps. Kreindler and Young (2014) provides sufficient conditions on payoffs and noise for evolution to be fast on arbitrary networks.

The methodological parts of our paper are related to the literature on stochastic stability with constant step size, where the dimension of the Markov process is held fixed as the
population grows. \footnote{Kaniovski and Young (1995) and Benaïm and Hirsch (1999) provide results connecting discrete-time and continuous-time limit dynamics in the context of fictitious-play style models, but the results themselves are not closely related because the approximation occurs in the $t \to \infty$ limit with the population size held fixed.} Benaïm and Weibull (2003) considers models where all agents have the same beliefs (which is why the state space doesn’t depend on the population size) and a random player from an $N$ agent population is chosen to revise his play at $t = \frac{1}{N}, \frac{2}{N}, \frac{3}{N}, \ldots$. The paper shows that the maximum deviation between the finite $N$ model and its continuous limit over a $T$ period horizon goes to zero as $N \to \infty$, and also that the probability of exiting the basin of attraction of an attractor of the continuous model by time $T$ goes to zero (exponentially fast) as the population size increases; this is related to our “leaving slowly” results. Roth and Sandholm (2013) develops more general results along these lines, and provides conditions that imply that play in the discrete-time model remains within $\epsilon$ of the dynamics of the continuous-time model for at least $T$ periods with probability that goes to one as $N$ grows. The results can provide an alternate method for establishing some of our applied results in cases where the state space remains the same as the population grows.

6 Conclusion

We defined a notion of fast convergence for evolutionary models, which refines the previous literature by requiring that some set $A$ is both reached quickly and left slowly. We then used Lyapunov functions to give sufficient conditions for fast convergence. One advantage of our approach is that it can apply to models with state spaces that do not have a finite-dimensional continuous time limit dynamic. At a conceptual level, our proofs separate the factors sufficient from fast convergence from technical conditions needed to assure a well-behaved approach to the mean field. We illustrated the use of our conditions in various examples without presenting quantitative results but the simulations of e.g. Ellison (1993) and Kreindler and Young (2013) make us optimistic that when convergence is fast in our sense it will be fast enough to be of practical importance.
References


Appendix

Proof of Proposition 2. To show that condition (1) holds for every open set $U$ containing $\bar{A}$, it suffices to show that this is true for every open $U$ with $\bar{A} \subset U \subset U_0$ and $U \neq \Delta$. Given this restriction, condition (4) always holds until the process has left $U$. Since $V$ is Lipschitz continuous and the increments of $X^N$ are bounded by $\frac{K}{N}$, there is a constant $\kappa$ such that $P_s(|V(X_N^N(\frac{1}{N})) - V(X_N^N(0))| \leq \frac{c}{N}) = 1$ for all $s \in \Omega_N$ and all $N$. Using this bound, Taylor’s formula implies that

$$E_s e^{\delta N[V(X_N^N(\frac{1}{N})) - V(X_N^N(0))]} = 1 + \delta NE_s \left[ V\left(X_N^N\left(\frac{1}{N}\right)\right) - V(X_N^N(0)) \right] + R_{N,s},$$

where $|R_{N,s}| \leq \frac{1}{2} \delta^2 \kappa^2 e^{\delta \kappa}$. Pick $\delta_0 > 0$ so that $\delta_0 \kappa^2 e^{\delta_0 \kappa} < c$. Let $Y^N(t) = \exp(\delta_0 NV(X_N^N(t)))$ and $Z^N(t) = Y^N(t \wedge \tau_N^N \wedge \tau_N^{\Delta\setminus U})$. Suppose $N \geq N_0$. Then, by (4), for every $s \in \phi_N^{-1}(U \setminus \bar{A})$,

$$E_s Y^N\left(\frac{1}{N}\right) = e^{\delta_0 NV(\phi_N(s))} E_s e^{\delta_0 N[V(X_N^N(\frac{1}{N})) - V(X_N^N(0))]} \leq E_s Y^N(0) \left(1 - \delta_0 c + \frac{1}{2} \delta_0^2 \kappa^2 e^{\delta_0 \kappa}\right) \leq E_s Y^N(0).$$

Thus, $\{Z^N(t)\}$ is a nonnegative supermartingale.

Let $v_1 = \max_{x \in \bar{A}} V(x)$, $v_3 = \min_{x \in \Delta \setminus U} V(x)$. By (3), $v_1 < v_3$. Let $v_2 \in (v_1, v_3)$ and $\gamma = \frac{1}{2} \delta_0 (v_3 - v_2)$. If $s \in \Omega_N$ and $V(\phi_N(s)) \leq v_2$, then by the maximal inequality (see e.g. Shiryaev (1996), page 493),

$$P_s(\tau_N^{\Delta\setminus U} < \tau_N^\bar{A}) \leq P_s \left( \sup_{t \geq 0} Z^N(t) \geq e^{\delta_0 N v_3} \right) \leq e^{-\delta_0 N v_3} E_s Z^N(0)$$

$$= e^{-\delta_0 N v_3} e^{\delta_0 N V(\phi_N(s))} \leq e^{\delta_0 N(v_2 - v_3)} = e^{-2\gamma N}.$$

Let $\sigma_{-1}^N = -1$. Define stopping times $\sigma_0^N \leq \sigma_1^N \leq \ldots$ by

$$\sigma_k^N = \inf \left\{ t \in \frac{1}{N} \mathbb{N}_0 : t > \sigma_{k-1}^N, X^N(t) \in \bar{A} \cup (\Delta \setminus U) \right\}, \quad k \in \mathbb{N}_0.$$

In view of (4), $P_s(V(X_N^N(\frac{1}{N})) < V(X_N^N(0))) > 0$ for all $s \in \phi_1^{-1}(U \setminus \bar{A})$. Thus, $P_s(\tau_N^\bar{A} < \infty) > 0$ for all $s \in \phi_N^{-1}(U \setminus \bar{A})$. This implies that $P_s(\tau_N^\bar{A} \wedge \tau_N^{\Delta\setminus U} < \infty) = 1$ for all $s \in \Omega_N$, see Durrett (1996), page 290. Hence $P_s(\sigma_k^N < \infty \text{ for all } k) = 1$.

Assume from now on that $N \geq N_0$ is so large that $v_1 + \frac{c}{N} \leq v_2$. Then for every $s \in \phi_N^{-1}(\bar{A})$, $P_s(V(X_N^N(\frac{1}{N})) \leq v_2) = 1$ and so, by (9)

$$P_s(X_N^N(\sigma_1^N) \in \bar{A}) = \sum_{\xi \in \Omega_N : V(\phi_N(\xi)) \leq v_2} P_\xi \left( X_N^N(\sigma_0^N) \in \bar{A} \right) P_s \left( S^N\left(\frac{1}{N}\right) = \xi \right) \geq 1 - e^{-2\gamma N}.$$
Hence, for \( s \in \phi_N^{-1}(\bar{A}) \) and \( k = 0, 1, \ldots, \)

\[
P_s \left( N\tau_{\Delta|U}^N > k \right) \geq P_s \left( X^N(\sigma_j^N) \in \bar{A}, \ 0 \leq j \leq k \right) \geq \left( 1 - e^{-2\gamma N} \right)^k.
\]

It follows that for all \( T \in [0, \infty) \),

\[
P_s \left( \tau_{\Delta|U}^N > T \right) = P_s \left( N\tau_{\Delta|U}^N > \lfloor NT \rfloor \right) \geq \left( 1 - e^{-2\gamma N} \right)^{NT} \geq \exp \left( -Te^{-\gamma N} \right),
\]

provided that \( N \) is also so large that \( 2Ne^{-\gamma N} \leq 1 \). Here \( \lfloor NT \rfloor \) denotes the largest integer \( \leq NT \). In the last step, it was used that \( 1 - u \geq e^{-2u} \) for \( u \in [0, \frac{1}{2}] \). \( \square \)

**Proof of Proposition 4.** Since \( \{X^N\} \) reaches \( B \) and \( A \cup B^c \) quickly, there exist \( N_1 \in \mathbb{N} \) and \( K < \infty \) so that for all \( N \geq N_1 \) and all \( s \in \Omega_N \), \( E_s\tau_B^N < K \) and \( E_s\tau_{A \cup B^c}^N < K \). Fix \( N \geq \max\{N_0, N_1\} \). Let \( \sigma_0 = 0 \) and for \( j = 0, 1, \ldots, \)

\[
\sigma_{2j+1} = \inf \{ t \geq \sigma_{2j} : X^N(t) \in B \},
\]
\[
\sigma_{2j+2} = \inf \{ t \geq \sigma_{2j+1} : X^N(t) \in A \cup B^c \}.
\]

Note that \( P_s(\sigma_j < \infty \text{ for all } j) = 1 \) for every \( s \in \Omega_N \). Let \( J = \inf \{ j \in \mathbb{N}_0 : X^N(\sigma_j) \in A \} \). Then, by Fubini,

\[
E_s\tau_A^N = E_s \sum_{j=1}^J (\sigma_j - \sigma_{j-1}) = \sum_{j=1}^\infty E_s[(\sigma_j - \sigma_{j-1})1_{\{J \geq j\}}].
\]

Let \( \mathcal{F}_t \) denote the \( \sigma \)-algebra generated by \( S^N(0), \ldots, S^N(t) \), let \( \mathcal{F} \) denote the \( \sigma \)-algebra generated by \( \cup_t \mathcal{F}_t \), and let \( \mathcal{F}_{\sigma_j} \) denote the \( \sigma \)-algebra up to time \( \sigma_j \), that is, \( \mathcal{F}_{\sigma_j} = \{ C \in \mathcal{F} : C \cap \{ \sigma_j = t \} \in \mathcal{F}_t \text{ for all } t \} \). For every \( j \geq 1 \),

\[
\{ J \geq j \} = \{ X^N(\sigma_0) \not\in A, \ldots, X^N(\sigma_{j-1}) \not\in A \}
\]
\[
= \{ S^N(\sigma_0) \not\in \phi_N^{-1}(A), \ldots, S^N(\sigma_{j-1}) \not\in \phi_N^{-1}(A) \},
\]

so that \( \{ J \geq j \} \in \mathcal{F}_{\sigma_{j-1}} \). Hence,

\[
E_s[(\sigma_j - \sigma_{j-1})1_{\{J \geq j\}}] = E_s[E[(\sigma_j - \sigma_{j-1})1_{\{J \geq j\}}|\mathcal{F}_{\sigma_{j-1}}]]
\]
\[
= E_s[1_{\{J \geq j\}}E[\sigma_j - \sigma_{j-1}|\mathcal{F}_{\sigma_{j-1}}]].
\]

By the strong Markov property,

\[
E[\sigma_j - \sigma_{j-1}|\mathcal{F}_{\sigma_{j-1}}] = \begin{cases} 
E_{S^N(\sigma_{j-1})\tau_B^N}^{\sigma_{j-1}} \leq K & \text{if } j \text{ is odd,} \\
E_{S^N(\sigma_{j-1})\tau_{A \cup B^c}^N}^{\sigma_{j-1}} \leq K & \text{if } j \text{ is even.}
\end{cases}
\]
Thus, $E_s \tau^N_A \leq K \sum_{j=1}^{\infty} P_s(J \geq j) \leq 2K \sum_{j=0}^{\infty} P_s(J \geq 2j+1)$. If $P_s(J \geq 2j+1) > 0$, then

$$P_s(J \geq 2j+1) = P_s(X^N(\sigma_k) \not\in A \text{ for } k = 0, \ldots, 2j) = \prod_{k=1}^{2j} P_s(X^N(\sigma_k) \not\in A \mid X^N(\sigma_k) \not\in A \text{ for } \kappa = 0, \ldots, k-1).$$

When $k$ is even, the conditional probability is at most $1 - c$ by (5). Hence $E_s \tau^N_A \leq 2K \sum_{j=0}^{\infty} (1-c)^j = 2K/c$. □

**Proof of Example 1.** If $x \in \Delta$ and $x_1 \leq \frac{1}{2}$, then

$$\sum_{i=2}^{m} e^{\beta \pi_i(x)} \leq \sum_{i=2}^{m} e^{\beta \alpha_2 x_1} \leq m - 2 + e^{\beta \alpha_2 (1-x_1)} \leq m - 2 + e^{\beta \alpha_2},$$

where the second inequality follows from the Schur-convexity of $\sum_{i=2}^{m} e^{\beta \alpha_2 x_1}$, see e.g. Marshall and Olkin (1979), page 64. Hence, $f_1(x) \geq h(x_1)$, where

$$h(x_1) = \frac{1}{1 + (m - 2 + e^{\beta \alpha_2}) e^{-\beta \alpha_1 x_1}}.$$  

Since $\beta \alpha_1 > 2(m - 2 + e^{\beta \alpha_2})/e$ and $e^u \geq eu$ for all $u \geq 0$, $e^{\beta \alpha_1 x_1} > 2(m - 2 + e^{\beta \alpha_2}) x_1$ for $x_1 > 0$. Consequently,

$$h(x_1) > \frac{1}{1 + \frac{1}{2x_1}} \geq x_1$$

for $0 < x_1 \leq \frac{1}{2}$. Since $h$ is continuous and $h(0) > 0$, it follows that there exists $c > 0$ so that $f_1(x) - x_1 \geq h(x_1) - x_1 \geq c$ for all $x \in \Delta$ with $x_1 \leq \frac{1}{2}$. The assertion follows from Proposition 5(a). □

**Proof of Example 2.** The result can again be shown by applying Proposition 5(a), but instead we apply Proposition 7 to also justify the claim at the end of Subsection 4.1. The probability $f_1(x, \beta)$ is uniquely determined by $\beta[\pi_1(x) - \pi_2(x)] = \psi(f_1(x, \beta))$, where $\psi(p) := c'(p) - c'(1-p)$, $0 < p < 1$. The function $\psi$ is continuous and strictly increasing, $\lim_{p \to 0^+} \psi(p) = -\infty$, and $\lim_{p \to 1^-} \psi(p) = \infty$. The inverse function $\psi^{-1}$ is therefore a continuous strictly increasing function on $\mathbb{R}$ with $\lim_{u \to -\infty} \psi^{-1}(u) = 0$ and $\lim_{u \to \infty} \psi^{-1}(u) = 1$. Let $\epsilon$ be a random variable that has $\psi^{-1}$ as its distribution function. Then the support of $\epsilon$ is $\mathbb{R}$, $P(\epsilon = 0) = 0$, and $f_1(x, \beta) = \psi^{-1}(\beta[\pi_1(x) - \pi_2(x)]) = P(\beta[\pi_1(x) - \pi_2(x)] \geq \epsilon)$. Since $\psi(1-p) = -\psi(p)$ for all $p$, $1 - \psi^{-1}(u) = \psi^{-1}(-u)$ for all $u$, which implies that $\epsilon$ and $-\epsilon$ have the same distribution. The assertion now follows from Proposition 7. □
Proof of Example 3. For \( x \in \Delta, 0 \leq \epsilon \leq 1 - \frac{1}{m} \), and \( 0 \leq y \leq 1 \) let
\[
q(x, \epsilon) = x_1(1 - \epsilon) + (1 - x_1) \frac{\epsilon}{m - 1}; \quad H(y) = 1 - (1 - y)^s.
\]
In state \( x \), the probability that a randomly sampled agent is thought to have played strategy 1 is \( q(x, \epsilon) \), so that in a sample of size \( s \), the probability that a revising agent chooses strategy 1 is \( f_1(x, \epsilon) = H(q(x, \epsilon)) \).

Obviously, \( H \) is strictly increasing, and \( (\partial/\partial\epsilon)q(x, \epsilon) = (1 - mx_1)/(m - 1) \). Hence, for every \( x \in \Delta, f_1(x, \epsilon) \) is strictly increasing in \( \epsilon \) if \( x_1 < \frac{1}{m} \), and \( f_1(x, \epsilon) \) is strictly decreasing in \( \epsilon \) if \( x_1 > \frac{1}{m} \).

Let \( \xi = (\xi_1, \ldots, \xi_m) \) be a point in \( \Delta \) with \( \xi_1 = 1 - (1 - \frac{1}{m})^s \). Then \( f_1(\xi, 1 - \frac{1}{m}) = \xi_1 \).

Since \( s \geq 2 \), \( \xi_1 > \frac{1}{m} \), so that \( f_1(\xi, \epsilon) \) is strictly decreasing in \( \epsilon \). Thus, \( f_1(\xi, \epsilon) - \xi_1 > 0 \) for every \( \epsilon \in (0, 1 - \frac{1}{m}) \). Also, if \( x \in \Delta \) and \( x_1 = 0 \), then \( f_1(x, \epsilon) > 0 \) for \( \epsilon > 0 \). Since \( H \) is concave, it follows that for every \( \epsilon \in (0, 1 - \frac{1}{m}) \), \( f_1(x, \epsilon) - x_1 > 0 \) for all \( x \in \Delta \) with \( x_1 \leq \xi_1 \).

In view of the continuity of \( f_1 \) it now follows from Proposition 5(a) that \( \{X^N\} \) has fast convergence to \( \{x : x_1 > \xi_1\} \). \( \square \)

The proof of Proposition 6 uses two lemmas. The first notes that in a 2 \( \times \) 2 game Proposition 4 implies that we have fast convergence if we can show that \( \{X^N\} \) quickly reaches some threshold and the dynamics are monotone from that point. The second provides the desired monotonicity result for the dynamics with sampling. In the following the state \( x \in [0, 1] \) is the fraction of agents using strategy 1 and \( g(x) \) is the probability that a revising agent chooses strategy 1.

Lemma 1. In the model described above let \( a \) and \( \xi \) be constants with \( 0 < a < \xi < 1 \) and suppose \( c > 0 \). Suppose that \( \{X^N\} \) reaches \( [a, 1] \) quickly and
\[
g(x) - x \geq c \text{ for all } x \in [a, \xi].
\]
Then \( \{X^N\} \) has fast convergence to \( (\xi, 1] \).

Proof of Lemma 1. Let \( A = (\xi, 1], B = [a, 1] \). It follows from Proposition 1 with \( V(x_1, x_2) = 1 - x_1 \) that \( \{X^N\} \) reaches \( A \cup B^c \) quickly. For \( x \in (B \setminus A) \cap \{0, \frac{1}{N}, \ldots, \frac{N}{N}\} \),
\[
\frac{P_x(X^N(\frac{1}{N}) = x + \frac{1}{N})}{P_x(X^N(\frac{1}{N}) = x - \frac{1}{N})} = \frac{(1 - x)g(x)}{x(1 - g(x))} \geq \frac{(1 - x)(x + c)}{x(1 - x - c)} \geq c_0,
\]
where \( c_0 := (1 + c)^2/(1 - c)^2 \). Hence, by the formula for absorption probabilities of birth and death chains, see e.g. Karlin and Taylor (1975), page 113,
\[
P_x(X^N(\tau_{A \cup B^c}) \in A) \geq \frac{1}{\sum_{k=0}^{\infty} c_0^k} = 1 - c_0^{-1} > 0.
\]

26
Thus, by Proposition 4, \( \{X^N\} \) reaches \( A \) quickly. That \( \{X^N\} \) leaves \( A \) slowly follows from Proposition 2 with \( V(x_1, x_2) = 1 - x_1 \) as the Lyapunov function. □

**Lemma 2.** Suppose \( g(0) > 0 \), \( g(\frac{k}{s}) \geq \frac{k}{s} \) for all \( k \leq \frac{s}{2} \), \( g(\frac{k}{s}) + g(1 - \frac{k}{s}) \geq 1 \) for all \( k = 0, \ldots, s \), and \( g(\frac{k}{s}) + g(1 - \frac{k}{s}) > 1 \) for some \( k \). Then \( \min\{f^{(s)}(x) - x : 0 \leq x \leq \frac{1}{2}\} > 0 \).

**Proof of Lemma 2.** Set \( \tilde{g}(x) = g(x) - x \). Then \( \tilde{g}(1 - \frac{k}{s}) \geq -\tilde{g}(\frac{k}{s}) \) for all \( k \). Hence for all \( x \in [0, 1] \),

\[
f^{(s)}(x) - x = \sum_{k \leq \frac{s}{2}} \tilde{g} \left( \frac{k}{s} \right) \left( \frac{s}{k} \right) x^k (1 - x)^{s-k} + \sum_{k < \frac{s}{2}} \tilde{g} \left( 1 - \frac{k}{s} \right) \left( \frac{s}{k} \right) x^{s-k} (1 - x)^k \]

\[
\geq \sum_{k < \frac{s}{2}} \tilde{g} \left( \frac{k}{s} \right) \left( \frac{s}{k} \right) \left[ x^k (1 - x)^{s-k} - x^{s-k} (1 - x)^k \right]. \tag{10}
\]

If \( \frac{k}{s} \), then \( \tilde{g}(\frac{k}{s}) \geq 0 \) and the term in square brackets is nonnegative for all \( 0 \leq x \leq \frac{1}{2} \). Thus for all \( 0 \leq x < \frac{1}{2} \),

\[
f^{(s)}(x) - x \geq g(0)[(1 - x)^s - x^s] > 0.
\]

Since \( g(\frac{k}{s}) + g(1 - \frac{k}{s}) > 1 \) for some \( k \), the inequality in (10) is strict for \( x = \frac{1}{2} \), and so \( f^{(s)}(\frac{1}{2}) - \frac{1}{2} > 0 \). □

**Proof of Proposition 6.** Suppose \( \{X^N\} \) reaches \( [x^*, 1] \) quickly. To apply Lemma 2 note first that if there existed some \( x_0 \in (0, x^*) \) with \( g(x_0) < x_0 \), then, since \( g \) is nondecreasing, for all \( x \in (\frac{1}{2}(g(x_0) + x_0), x_0) \),

\[
g(x) - x \leq g(x_0) - \frac{g(x_0) + x_0}{2} = \frac{g(x_0) - x_0}{2} < 0.
\]

By the remark after Proposition 5, it would follow that \( \{X^N\} \) does not reach \( [x^*, 1] \) quickly. Thus \( g(x) \geq x \) for all \( 0 \leq x \leq x^* \). As \( g(x^*) \geq \frac{1}{2} \) and \( g \) is strictly increasing, there exists \( \delta > 0 \) such that \( g(x) \geq x + \delta \) for all \( x \in [x^*, \frac{1}{2} + \delta] \). For all \( x \in [0, \frac{1}{2}] \),

\[
\frac{1}{2} + x > x^* + \min(x, x^*), \quad \frac{1}{2} - x \geq x^* - \min(x, x^*),
\]

and so

\[
g \left( \frac{1}{2} + x \right) + g \left( \frac{1}{2} - x \right) > g(x^* + \min(x, x^*)) + g(x^* - \min(x, x^*)) \geq 1.
\]

In particular, \( g(0) > 0 \). It now follows from Lemma 2 that \( \min\{f^{(s)}(x) - x : 0 \leq x \leq \frac{1}{2}\} > 0 \) for every \( s \).

27
As $g$ is nondecreasing, so is each $f^{(s)}$, and $\lim_{s \to \infty} f^{(s)}(1/2) = 1/2 [g(1/2^-) + g(1/2^+)] > 1/2$ (see Lorentz (1986), pages 23 and 27.) Since $f^{(s)}(1/2) > 1/2$ for every $s$, it follows that there exists $\xi \in (1/2, 1/2 + \delta)$ so that $f^{(s)}(1/2) > \xi$ for every $s$. Hence, if $x \in [1/2, \xi]$, then $f^{(s)}(x) - x \geq f^{(s)}(1/2) - \xi > 0$. Consequently,

$$\min_{0 \leq x \leq \xi} f^{(s)}(x) - x > 0 \quad \text{and} \quad \inf_{x^* \leq x \leq \xi} g(x) - x > 0.$$  

Therefore, by Proposition 5(a), $\{X^N_s\}$ has fast convergence to $(\xi, 1]$ and by Lemma 1, $\{X^N\}$ has fast convergence to the same set. □

**Proof of Proposition 7.** Write $h(u)$ for $P[u \geq \epsilon]$. Note that the restrictions on the distribution of $\epsilon$ imply that $h$ is strictly increasing and satisfies

$$h(u) + h(-u) \geq 1 \text{ for all } u, \quad \lim_{u \to -\infty} h(u) < x^*, \quad \lim_{u \to 0-} h(u) > x^*.$$  

For $\beta > 0$ let $G(\beta) = \inf\{g(x, \beta) - x : 0 \leq x \leq x^*\}$. Let $\beta^* = \sup\{\beta > 0 : G(\beta) \geq 0\}$. Since $\lim_{u \to 0-} h(u) > x^*$, $G(\beta) > 0$ for some small $\beta > 0$, and so $\beta^* > 0$. If $x \in (\lim_{u \to -\infty} h(u), x^*)$, then $\lim_{\beta \to \infty} g(x, \beta) - x < 0$. Thus, $\beta^* < \infty$. To prove claims (i) and (ii) it will suffice to show that for some $\beta^*_s > \beta^*$ three results hold:

(a) if $0 < \beta < \beta^*$, $\{X^N\}$ has fast convergence to $(1/2, 1]$;

(b) if $\beta > \beta^*$, $\{X^N\}$ does not reach $[x^*, 1]$ quickly;

(c) if $0 < \beta < \beta^*_s$, $\{X^N_s\}$ has fast convergence to $(1/2, 1]$.

(a) Let $0 < \beta < \beta^*$. Then there exists $\beta' \in (\beta, \beta^*)$ with $G(\beta') \geq 0$. As $g(x^*, \beta) = h(0) \geq 1/2$ and $\lim_{u \to 0-} h(u) > x^*$, there exists $\delta > 0$ such that $g(x, \beta) - x \geq \delta$ for all $x \in [x^* - \delta, x^*]$. If $x \in [0, x^* - \delta]$, then $x' := \frac{\beta}{\beta'}(x - x^*) + x^* \in [0, x^*]$ and $\beta[\pi_1(x) - \pi_2(x)] = \beta'[\pi_1(x') - \pi_2(x')]$, so that

$$g(x, \beta) - x = g(x', \beta') - x' + (x^* - x) \left(1 - \frac{\beta}{\beta'}\right) \geq G(\beta') + \delta \left(1 - \frac{\beta}{\beta'}\right).$$  

Hence $g(x, \beta) - x \geq \delta(1 - \beta/\beta')$ for all $x \in [0, x^*]$. Since $g(x, \beta)$ is strictly increasing in $x$ and $g(x^*, \beta) \geq 1/2$, it follows that $\inf_{0 \leq x \leq 1/2} g(x, \beta) - x > 0$ and so, by Proposition 5(a), $\{X^N\}$ has fast convergence to $(1/2, 1]$.

(b) Let $\beta > \beta^*$. Then $G(\beta) < 0$, so that for some $x_0 \in (0, x^*)$, $\delta := x_0 - g(x_0, \beta) > 0$. Since $g(x, \beta)$ is increasing in $x$, $g(x, \beta) - x \leq -\frac{\delta}{2}$ for all $x \in [x_0 - \delta/2, x_0]$. Hence, by the remark after Proposition 5, $\{X^N\}$ does not reach $[x^*, 1]$ quickly.

28
(c) To prove the assertion about \( \{X^N_s\} \) we first show that for \( s \geq 2 \),
\[
\inf_{\frac{1}{2} \leq x \leq \frac{1}{2}} \; g(x, \beta) - x \geq 1 - \frac{\beta}{\beta^*} \quad \text{if } \beta \geq \beta^* \text{ and } (s - 2)\beta \leq s\beta^*. \tag{11}
\]

Suppose \( \beta \) satisfies both conditions. If \( x^* \leq x \leq \frac{1}{2} \), then \( g(x, \beta) - x \geq g(x^*, \beta) - \frac{1}{2} \geq 0 \geq 1 - \frac{\beta}{\beta^*} \). Suppose now \( \frac{1}{2} \leq x \leq x^* \). Let \( \{\beta_n\} \) be a sequence with \( G(\beta_n) \geq 0 \) for all \( n \) and \( \beta_n \to \beta^* \). Let \( x_n = \frac{\beta}{\beta_n}(x - x^*) + x^* \). Then \( x_n \leq x^* \) for all \( n \) and \( \lim_{n \to \infty} x_n > 0 \). Thus, for \( n \) sufficiently large, \( x_n \in [0, x^*] \), so that \( g(x_n, \beta_n) - x_n \geq G(\beta_n) \geq 0 \) and
\[
g(x, \beta) - x = g(x_n, \beta_n) - x_n + (x^* - x) \left( 1 - \frac{\beta}{\beta_n} \right) \geq 1 - \frac{\beta}{\beta_n}.
\]

Letting \( n \to \infty \) completes the proof of (11).

Since \( x^* < \frac{1}{2} \), \( \pi_1(1 - x) - \pi_2(1 - x) > \pi_2(x) - \pi_1(x) \), and so,
\[
g(x, \beta) + g(1 - x, \beta) > h(\beta[\pi_1(x) - \pi_2(x)]) + h(\beta[\pi_2(x) - \pi_1(x)]) \geq 1
\]
for all \( x \in [0, 1] \) and \( \beta > 0 \). Consequently,
\[
f^{(s)}(x, \beta) - x = \sum_{k=0}^{s} \left( g \left( \frac{k}{s}, \beta \right) - \frac{k}{s} \right) p_{s,k}(x)
\geq g(0, \beta)(1 - x)^s + [g(1, \beta) - 1]x^s
+ \sum_{1 \leq k < \frac{s}{2}} \left( g \left( \frac{k}{s}, \beta \right) - \frac{k}{s} \right) [p_{s,k}(x) - p_{s,s-k}(x)],
\]
where \( p_{s,k}(x) = \binom{s}{k} x^k (1 - x)^{s-k} \). Recall from (i) that \( g(\frac{k}{s}, \beta) \geq \frac{k}{s} \) if \( k \leq \frac{s}{2} \) and \( \beta < \beta^* \). Moreover, \( 0 \leq p_{s,k}(x) - p_{s,s-k}(x) \leq 1 \) if \( k \leq \frac{s}{2} \) and \( x \in [0, \frac{1}{2}] \). It now follows by (11) that for every \( \beta > 0 \) with \( (s - 2)\beta \leq s\beta^* \),
\[
\min_{x \in [0, \frac{1}{2}]} f^{(s)}(x, \beta) - x \geq [g(0, \beta) + g(1, \beta) - 1] \left( \frac{1}{2} \right)^s - \frac{s}{2} \left( 1 - \frac{\beta}{\beta^*} \right)^-, \]
where \( u^- = - \min(u, 0) \). Since \( \pi_1(1) - \pi_2(1) > \pi_2(0) - \pi_1(0) \), \( g(0, \beta) + g(1, \beta) > 1 \) for every \( \beta > 0 \), and
\[
\liminf_{\beta \to \beta^*} g(0, \beta) + g(1, \beta) \geq g(0, \beta^*) + g(1, \beta^*) > 1.
\]
Therefore, for every \( s \in \mathbb{N} \) there exists \( \beta^*_s > \beta^* \) so that if \( \beta < \beta^*_s \), then \( \inf_{x \in [0, \frac{1}{2}]} f^{(s)}(x, \beta) - x > 0 \). By Proposition 5(a), if \( \beta < \beta^*_s \), then \( \{X^N_s\} \) has fast convergence to \( (\frac{1}{2}, 1] \). □

**Proof of claim in footnote 12.** Assume the payoff functions \( \pi_i \) are as in Proposition 7 and the choice rule \( g(x, \beta) \) is generated by a perturbed utility function with cost function
$c(p) = p^2$. To show that the conclusion of Proposition 7 does not hold we show that for every parameter $\beta > 0$ either $\{X^N\}$ and $\{X_s^N\}$ have fast convergence to $(\frac{1}{2}, 1]$ or neither system has.

For the present cost function $c$, $g(x, \beta) = h(\beta[\pi_1(x) - \pi_2(x)])$, where

$$h(u) = \begin{cases} 
0, & u \leq -2 \\
\frac{u+2}{4}, & -2 < u < 2, \\
1, & u \geq 2.
\end{cases}$$

Let $\beta^* = 2/\left[\pi_2(0) - \pi_1(0)\right]$. If $\beta \geq \beta^*$, then $g(0, \beta) = 0$ and so $0$ is an absorbing state of $X^N$ and $X_s^N$ for every $N$. In particular, neither $\{X^N\}$ nor $\{X_s^N\}$ has fast convergence to $(\frac{1}{2}, 1]$. □

**Proof of Example 5.** Let $s$ be any state in which a fraction $x_1$ of the players play strategy 1. The expected change in the fraction of players using strategy 1 will be

$$E(X_1(t + \frac{1}{N}) - x_1 | Z(t) = z) = \frac{1}{N} \left[ y(1 - \epsilon) + (1 - y)\epsilon - x_1 \right],$$

where $y$ is the expected fraction of players who have strategy 1 as a best response in state $s$. Note that the RHS can be reorganized as

$$y(1 - \epsilon) + (1 - y)\epsilon - x_1 = y(1 - 2\epsilon) + \epsilon - x_1$$

$$= (y - x_1)(1 - 2\epsilon) + \epsilon(1 - 2x_1)$$

Any state $s$ with fraction $x_1$ players playing strategy 1 will have fraction $r$ players with two neighbors playing strategy 1, fraction $2(x_1 - r)$ with one neighbor playing strategy 1, and fraction $(1 - 2x_1 + \ell)$ with no neighbor playing strategy 1 for some $r \in [0, x_1]$. The value of $y$ depends on the state $s$ only through $x_1$ and $r$. Because players with two neighbors playing strategy 1 will always have strategy 1 as their best response, those with one neighbor playing strategy 1 will have strategy 1 as their best response if at least one player they randomly sample uses strategy 1, and those with no neighbors playing strategy 1 must have both
players in their sample using strategy 1, we have \( y = r + 2(x_1 - r)(2x_1 - x_1^2) + (1 - 2x_1 + r)x_1^2 \).

Writing \( x_1 = r + (x_1 - r) \) and collecting terms gives

\[
y - x_1 = (x_1 - r)(-1 + 4x_1 - 2x_1^2) + ((1 - x_1) - (x_1 - r))x_1^2
\]

\[
= (x_1 - r)(-1 + 4x_1 - 3x_1^2) + (1 - x_1)x_1^2
\]

Plugging back into the formula for the change in \( x_1 \) gives

\[
NE(X_1(t + \frac{1}{N}) - x_1|Z(t) = z) = ((x_1 - r)(-1 + 4x_1 - 3x_1^2) + (1 - x_1)x_1^2) (1 - 2\epsilon) + \epsilon(1 - 2x_1),
\]

We show that the RHS can be bounded below by some positive constant \( c \) by considering three cases.

First, for \( x_1 \in [0, \frac{1}{3}] \) the quadratic \((-1 + 4x_1 - 3x_1^2) = (3x_1 - 1)(1 - x_1)\) is negative. Hence, the RHS is minimized for \( r = 0 \) in which case it is equal to

\[
(-x_1 + 5x_1^2 - 4x_1^3)(1 - 2\epsilon) + (1 - 2x_1)\epsilon
\]

\[
= (-x_1 + 5x_1^2 - 4x_1^3) + (1 - 10x_1^2 + 8x_1^3)\epsilon
\]

The polynomial \(-x_1 + 5x_1^2 - 4x_1^3\) is only negative if \( x_1 \) is additionally less than \( \frac{1}{4} \) and \( 1 - 10x_1 + 8x_1^3 \) is positive in this case, so the RHS will be bounded away from zero for \( x_1 \in [0, \frac{1}{3}] \) if \( \epsilon \) is chosen to be greater than

\[
sup_{x_1 \leq \frac{1}{4}} \frac{-x_1 + 5x_1^2 - 4x_1^3}{1 - 10x_1^2 + 8x_1^3}.
\]

Evaluating this numerically shows that choosing \( \epsilon = 0.065 \) suffices.

Second for \( x_1 \in [\frac{1}{3}, \frac{1}{2}] \) the expected change in \( x_1 \) is minimized for \( r = x_1 \) in which case it is simply \((1 - x_1)x_1^2(1 - 2\epsilon) + \epsilon(1 - 2x_1)\). This is obviously bounded away from zero for all \( \epsilon < \frac{1}{2} \).

Finally, for \( x_1 \in [\frac{1}{2}, 1 - \ell\epsilon] \) the minimum again occurs for \( r = x_1 \) and the value is again \((1 - x_1)x_1^2(1 - 2\epsilon) + \epsilon(1 - 2x_1)\) which expands as \( x_1^2 - x_1^3 + (1 - 2x_1 - 2x_1^2 + 2x_1^3)\epsilon \). The first term is positive and the second negative, so for each \( x_1 \) the expression is minimized by choosing \( \epsilon \) as large as possible given \( x_1 \): \( \epsilon = (1 - x_1)/\ell \). Factoring out the \( 1 - x_1 \) we find that this holds for all \( x_1 \) in the range if \( \ell \) is chosen to be greater than \( sup_{x_1 > 0.5} \frac{-2x_1^2 + 2x_1^2 + 2x_1 - 1}{x^2} \).

The maximum is about 1.42. □