

# Imitation and Experimentation in Changing Contests<sup>1</sup>

Francesco Squintani

*Department of Economics, University of Rochester, Rochester, New York*  
sqnt@troi.cc.rochester.edu

and

Juuso Välimäki

*Department of Economics, University of Southampton, Southampton, United Kingdom*  
juuso.valimaki@soton.ac.uk

Received May 8, 2000; final version received August 3, 2001

This paper analyzes the equilibrium play of individuals that are randomly matched to play a contest where the dominant action changes over time. Under myopic decision making, players adopt imitation strategies similar to those observed in evolutionary models with sampling from past play in the population. If the players are patient, equilibrium strategies display elements of experimentation in addition to imitation. If the changes in the environment are infrequent enough, these strategies succeed in coordinating almost all of the players on the dominant action almost all of the time. The myopic rules, on the other hand, result in miscoordination for a positive fraction of time. *Journal of Economic Literature* Classification Numbers: C73, D83, D84. © 2001 Academic Press

## 1. INTRODUCTION

When economic agents observe outcome samples from the past play in the population, they can react to the observations in two possible ways. They may decide to *imitate* the players who are using the most effective actions or to *experiment* with an alternative strategy that they did not observe in the population sample. When the environment is stable, models in evolutionary game theory predict that under mild regularity conditions, myopic players adopt imitative behavior and select the dominant action whenever it exists. This paper considers an equilibrium model where the environment changes from period to period, and dominant actions become

<sup>1</sup> We wish to thank Eddie Dekel, Jeff Ely, George Mailath, Wolfgang Pesendorfer, Karl Schlag, Lones Smith and Asher Wolinsky for helpful conversations. All remaining errors are ours.



dominated at random times. With myopic players, imitation strategies are still selected in equilibrium, but the players are not coordinated on the dominant action all the time. However, in a model with forward-looking players, we show that sampling from population play yields an equilibrium where both imitation and experimentation are present. Even though experimentation provides a public good in the model, the equilibrium rate of experimentation is sufficiently high to coordinate almost all of the players on the dominant action almost all the time if the changes in the environment are infrequent enough.

The model we analyze has two states of nature and two actions. In the first state, the first action is dominant, in the second state, it is dominated. In order to represent information transmission through sampling from the population play, we imagine a continuum of identical players matched according to a Poisson arrival process. As is customary in the evolutionary game theory literature, we are interested in the relative payoff comparison between individuals, rather than the absolute payoffs received by a single player. In line with that purpose, we assume that the players are matched to play a zero sum game with the following property. Whenever the players choose the same action, the game ends in a draw regardless of the true state, so that the match is not informative on the state of nature. If, on the other hand, a player wins by playing, say, the first action, then he or she (and his or her opponent) can deduce the true state of the world at the moment of the match. The state changes according to a stationary Markov transition process, independent of any actions taken in the game.

We consider first the case in which players observe the entire history of play and maximize their myopic utility. The equilibrium in this game takes a form familiar in evolutionary game theory: players adopt purely imitative strategies where all players choose the same action as in the previous period until a loss is observed. Our main result in this context is that under these imitation dynamics, the population play is not responsive to state changes. In fact, while the population shares of players choosing either strategy take all the values in the open unit interval infinitely often, the population fraction of players choosing, say, the first action crosses any fixed value in the unit interval very infrequently in comparison to the frequency of state changes. In other words, most of the state changes do not affect the play of most of the players.

In the second model, we introduce forward-looking behavior and assume that all players maximize their expected stream of future utilities. For simplicity we assume that players retain only single period histories and hence they condition only on the outcome in the previous match.<sup>2</sup> It is not

<sup>2</sup>We will show that this boundedly-rational strategy yields almost the same payoff as fully-rational strategies for our case of interest.

hard to see that the symmetric adoption of purely imitative strategies cannot constitute an equilibrium for this game. If almost all players in the population are playing a fixed action regardless of the true state of nature, then it is optimal for an individual player to experiment, i.e., to choose an action different from the previous action following a draw. To see this, notice that the losses from an experiment last for a single period. The choice of a dominated action, in fact, results almost certainly in the detection of the true state in the next period, and hence the player will revert to his or her original play. The payoff implications of a successful experiment, however, persist for more than a single period. The benefits from an experiment accumulate until the next state change. If the state changes are infrequent enough, then the benefits outweigh the losses, and the symmetric use of imitation strategies cannot be an equilibrium.

We show that the model with forward-looking players has a symmetric (and stationary) mixed-strategy equilibrium where all the players randomize with the same probability following the observation of a draw in the previous match. The main result of the paper is the characterization of these equilibria for infrequent state changes. In particular, it is shown that the fraction of time that any fixed population share spends on the dominated action converges to zero as state changes become infrequent. In other words, almost all of the players choose the dominant action almost all of the time, as the state changes become rare. A consequence of this result is that, with infrequent state changes, it would not be in any given player's self interest to sample additional past observation at a positive cost.

The techniques that we develop for the analysis might be of use in other contexts such as search models in a changing economic environment. Between the state changes, aggregate play in this model is deterministic by the law of large numbers. When the state changes, the law of motion changes for the aggregate population. The resulting compound stochastic process is an example of a piecewise-deterministic process as described in Davis [8]. The ergodic theory of these processes is quite simple, and we can make repeated use of renewal theory.

The paper is organized as follows. Section 2 presents the literature review. Section 3 introduces the model. Section 4 analyzes myopic players. Section 5 contains the equilibrium analysis for the case of forward-looking players. Section 6 concludes, and the proofs are collected in the Appendix.

## 2. RELATED LITERATURE

This paper is connected to three strands of literature. In the literature on herding, Ellison and Fudenberg [12] identify conditions under which

players will select the correct action given the state of the world, when sampling from the population play and adopting a rule where the individuals may change their action only if they sample some players taking a better alternative. Banerjee and Fudenberg [2] allow players to adopt fully rational decision rules. They show that if individuals sample from the population in a proportional fashion and signals are informative enough to outweigh the prior, then at the unique stationary outcome all agents make the correct choice. To our knowledge, the current paper is the first to study experimentation and social learning in a changing environment with forward-looking agents. Smith and Sorensen [21] explicitly introduce forward-looking behavior in a fixed environment and show that the set of stationary cascades shrinks as individuals become more patient. Moscarini *et al.* [17] analyze a social learning model in a changing world with myopic players.

The implications of sampling from population play have been studied extensively in the evolutionary game literature. Boylan [5] identifies matching schemes that allow the approximation of the stochastic population evolution by means of a dynamic system. Nachbar [16], Friedman [13], and Samuelson and Zhang [19] independently introduce payoff-monotonic dynamics and show that in continuous time, iterated strictly dominated strategies will be extinct in the long-run population if the initial population play has full support (see also Dekel and Scotchmer [9], Cabrales and Sobel [7], Bjornestedt [3], and Hofbauer and Weibull [14]). Specific characterizations of payoff-monotonic dynamics have then been derived in models of learning by sampling the population play by Bjornestedt [3], Bjornestedt and Weibull [4], Schlag [20], and Borgers and Sarin [6].

Models of experimentation in a changing world were treated in the single-agent case by Rustichini and Wolinsky [18] and by Keller and Rady [15] in a setting where a monopolist chooses between a sure action and an uncertain alternative whose value changes randomly over time. They show that patient players will take the optimal action almost all of the time if the state changes are infrequent enough. In our model, the forward looking optimal experimentation aspect of these models is combined with the effects of social learning and imitation.

### 3. THE MODEL

A continuum population of players indexed by the points in the unit interval are matched according to a Poisson process with parameter  $\mu$  to play one of two symmetric  $2 \times 2$  contests,  $G_1$  or  $G_2$ . In other words, the probability that player  $j \in [0, 1]$  is matched to play within the time interval

$(t, t + \Delta t)$  is  $\mu \Delta t$  for  $\Delta t$  small. The two possible payoff matrices (for the row player)  $G_1$  and  $G_2$  are given by:<sup>3</sup>

$G_1$	$a_1$	$a_2$
$a_1$	0	1
$a_2$	-1	0
$G_2$	$a_1$	$a_2$
$a_1$	0	-1
$a_2$	1	0

Note that action  $a_i$  is strictly dominant in game  $G_i$  for  $i = 1, 2$ .<sup>4</sup>

Let  $a^j(t)$  denote the action that player  $j$  would choose if matched at instant  $t$ . Denote the set of player  $j$ 's opponents in period  $t$  by  $j(t) \in [0, 1] \cup \emptyset$ , where  $j(t) = \emptyset$  if  $j$  is not matched in period  $t$ . Define the function  $m^j(t) \equiv \sup\{m < t \mid j(m) \neq \emptyset\}$ . Notice that because of Poisson matching,  $\Pr\{m^j(t) < t\} = 1$ , and  $m^j(t)$  can be interpreted as the last time before  $t$  in which  $j$  was matched.<sup>5</sup>

Denote the event that game  $G_i$  is played in period  $t$  by  $\{\omega(t) = \omega_i\}$ . The state space describing uncertainty about the game in period  $t$  is then given by  $\Omega = \{\omega_1, \omega_2\}$ . We assume that the state changes are also governed by a Poisson process with parameter  $\lambda$ .<sup>6</sup> Let  $\{\tau_k\}_{k=1}^\infty$  be the random sequence of state changes, where  $\tau_k$  is the physical instant of the  $k$ th state switch. Formally, we set  $\tau_0 = 0$ , and we iteratively define  $\tau_k = \inf\{t > \tau_{k-1} \mid \omega(t) \neq \omega(\tau_{k-1})\}$  for any  $k > 0$ . Each state duration  $(\tau_k, \tau_{k+1})$  is called an  $\omega_i$ -duration if  $\omega(t) = \omega_i$  for  $t \in (\tau_k, \tau_{k+1})$ . The evolution of play in the population is governed by the strategies of the players and the random state changes.

<sup>3</sup> The normalization to unit gains and losses off the diagonal is made for convenience. The main results of the paper would go through in the more general case as well.

<sup>4</sup> If the players are not sure of which game  $G_i$  they are playing, they can tell the two games apart conditional on observing an outcome off the main diagonal. A diagonal outcome does not help the players in distinguishing between the two games. This simple specification allows us to focus our attention on the informational content of relative-payoff comparison among individuals and to rule out any informational content of the absolute value of a player's payoff.

<sup>5</sup> Since the payoffs are defined as expectations over the matching probabilities and other variables, we can assign any behavior to the zero probability events where  $m^j(t) = t$ , without changing payoffs.

<sup>6</sup> Alternatively, we could suppose that the state durations are drawn independently from a known distribution,  $F_i(T)$ , for state  $\omega_i$ . In other words, if there is a state change at instant  $t$  to state  $\omega_i$ , then  $\Pr(\omega(s) = \omega_i \text{ for } t < s < u) = F_i(u-t)$ .

## 4. MYOPIC OPTIMIZATION

To set a benchmark that is comparable to previous contributions in evolutionary game theory, we assume in this section that each player maximizes his or her payoff in a myopic fashion.

We first define the history observed by player  $j$ . Let  $\mathbf{t}$  be the vector of previous matching times of  $j$ . The vector of actions chosen by  $j$  in the previous matches is denoted by  $\mathbf{a}^j$ , and the actions taken by  $j$ 's opponents are denoted by  $\mathbf{a}^{j(0)}$ . Let  $\mathbf{u}_j(\mathbf{a}^j, \mathbf{a}^{j(0)})$  denote the vector of realized payoffs. The history observable to player  $j$  at time  $t$  is  $h^j(t) = (\mathbf{a}^j, \mathbf{a}^{j(0)}, \mathbf{t}, \mathbf{u}_j(\mathbf{a}^j, \mathbf{a}^{j(0)}), t)$ , where the last component underlines that strategies may depend on calendar time.

A pure (behavior) strategy of an arbitrary player  $j$  at instant  $t$  is then  $\sigma^j: h^j(t) \rightarrow \{a_1, a_2\}$ , and its mixed extension is  $\sigma^j: h^j(t) \rightarrow \Delta\{a_1, a_2\}$ , where  $\sigma_i^j(h^j(t))$  denotes the probability that player  $j$  takes action  $a_i$ . Denoting by  $a(t)$  the random action of a player from the population at time  $t$ , player  $j$  prefers action  $a_1$  to action  $a_2$  at time  $t$  if

$$E[u(a_1, a(t), \omega(t)) | h^j(t)] \geq E[u(a_2, a(t), \omega(t)) | h^j(t)].$$

The notion of equilibrium in this game is defined in the standard fashion.

**DEFINITION 4.1.** An equilibrium is a collection of strategies  $(\sigma^j)_{j \in [0, 1]}$  such that for all  $j$  and all  $h^j(t)$ ,  $\sigma_i^j(h^j(t)) > 0$  only if

$$E[u(a_i, a(t), \omega(t)) | h^j(t)] \geq E[u(a_j, a(t), \omega(t)) | h^j(t)]$$

for all pairs of actions  $(a_i, a_j)$ .

For any action  $a_i$ , we introduce the following event on the set of all possible histories of player  $j$ :

$$L_i^j(t) = \{h^j(t): a^j(m^j(t)) = a_i, u_j(a^j(m^j(t)), a^{j(m)}(m^j(t))) = -1\}.$$

In words,  $L_i^j(t)$  stands for the event where player  $j$  lost in the last match by playing action  $a_i$ . We denote as *pure imitation rule* the choice of maintaining the same action as the previous match, unless defeated by the opponent. Formally we say that player  $j$  adopts the pure imitation rule at time  $t$  if, for any action  $a_i$ ,

$$\sigma_i^j(h^j(t)) = 0 \quad \text{if } h^j(t) \in L_i^j(t), \quad \text{and} \quad \sigma_i^j(h^j(t)) = 1 \quad \text{otherwise.}$$

The first proposition shows that imitation is the unique equilibrium behavior in the model.

**PROPOSITION 4.1.** *For all  $\lambda$  and  $\mu$ , the unique equilibrium is such that each player  $j$  adopts the pure imitation rule at all times  $t$ .*

The intuition is simple. Since the gain from a successful experiment is equal to the loss from an unsuccessful experiment, players will experiment only when assessing that the probability that the state has changed is at least one half. This is never the case under Poisson arrivals, because the probability that no state change has occurred dominates the probability that exactly one state change has occurred, and the probability that exactly  $2k$  state changes have occurred dominates the probability that  $2k+1$  state changes have occurred, for any  $k > 0$ .

It is important to notice that if  $G_i$  is the game played in all  $t$ , the population dynamics induced by the imitation rule leads to an asymptotic steady state where all players correctly assess the state of the world  $\omega_i$  and play action  $a_i$ . The following analysis shows that in changing environments, such assessments are often incorrect.

Denote the population fraction of players using  $a_1$  in period  $t$  by  $x(t)$ . By using the law of large numbers, we have

$$x(t) = \Pr\{a(t) = a_1\}$$

for a randomly picked player in  $t$ . To obtain a characterization of the rates of change of the actions in the population, we need to make a distinction according to the state of nature that prevails at  $t$ . Since the state changes according to a Poisson process, the time derivatives of the population fractions exist almost everywhere. As long as the state remains  $\omega_1$ , the law of motion for  $x(t)$  is given (almost everywhere) by:

$$\dot{x}(t) = \mu(1 - x(t)) x(t). \quad (1)$$

Of all the matched players (that have instantaneous flow rate of  $\mu$ ), only those playing  $a_2$  (fraction  $(1 - x_t)$ ) that are matched with players playing  $a_1$  (fraction  $x(t)$ ) adjust their behavior with positive probability. The solution to this differential equation yields the population share  $x(t)$  given an initial condition  $x(0)$ :

$$x(t) = \frac{\frac{x(0)}{1-x(0)} e^{\frac{\mu}{2}t}}{\frac{x(0)}{1-x(0)} e^{\frac{\mu}{2}t} + e^{-\frac{\mu}{2}t}} = \frac{1}{1 + \frac{1-x(0)}{x(0)} e^{-\mu t}}. \quad (2)$$

The dynamics for the population fraction playing  $a_2$  follows immediately from the constant population assumption. A similar derivation can be done for state  $\omega_2$  to yield:

$$x(t) = \frac{1}{1 + \frac{1-x(0)}{x(0)} e^{\mu t}}. \quad (3)$$

The main task in this section is to patch these two dynamics together to yield the overall population dynamics in the changing environment.

Before presenting the formal treatment, it may be useful to give an informal outline of the analysis. The first observation is that the closed forms (2) and (3) identify the movements of a variable along the same logistic curve, respectively from left to right and from right to left, at the same speed (controlled by  $\mu$ ). Thus equal durations in opposite states offset each other, and we can summarize the ergodic behavior of the variable by looking at the cumulative difference in durations. Consider three consecutive state changes: the expected time before the second change equals the expected time between the second and the third change. Thus the expected difference over two consecutive durations is equal to zero. Summing up differences over pairs of durations, one obtains a symmetric random walk, a recurrent process characterized by fluctuations that are both very wide and very slow (in a sense specified below). Very wide fluctuations in the time coordinate mean that the variable will spend almost all time arbitrarily close to the extremes of the interval  $[0, 1]$ . Very slow fluctuations mean that the variable will follow almost no state change. Thus the population will coordinate on the dominated action almost half of the time.

The formal treatment is as follows. Consider first the following limit

$$x_A = \lim_{T \rightarrow \infty} \frac{\int_0^T I_A(x(t)) dt}{T}, \quad \text{for } A \subset (0, 1),$$

where  $I_A(x(t)) = 1$  if  $x(t) \in A$  and  $I_A(x(t)) = 0$  if  $x(t) \notin A$ . The limit  $x_A$  measures asymptotically the fraction of time that  $x(t)$  spends in an arbitrary set  $A$ . The next lemma shows that this limit is 0 for all closed  $A$ .

**LEMMA 4.1.** *For any  $\lambda$  and  $\mu$ ,  $x_A = 0$  for all closed  $A \subset (0, 1)$ .*

We need the following definition to make precise the notion that the play in the population as described by  $x(t)$  is not very sensitive to state changes.

**DEFINITION 4.2.** The  $\omega_1$ -duration  $(\tau_k, \tau_{k+1})$  is  $\epsilon$ -undetected if  $x(t) \leq \epsilon$  for  $\tau_k \leq t \leq \tau_{k+1}$ . The  $\omega_2$ -duration  $(\tau_l, \tau_{l+1})$  is  $\epsilon$ -undetected if  $x(t) \geq 1 - \epsilon$  for  $\tau_l \leq t \leq \tau_{l+1}$ .



In words, at most fraction  $\epsilon$  of players in the population play the dominant action for the state duration in question. The previous lemma can be used to prove the following proposition.

**PROPOSITION 4.2.** *For all  $\lambda$  and  $\mu$ , the limiting frequency of the state durations that are  $\epsilon$ -undetected is  $\frac{1}{2}$  for all  $\epsilon > 0$ .*

An alternative statement would be that in the long run, the population play reacts to only a negligible subset of actual state changes, and as a result, the majority of the population play a constant action for a time interval that is by far longer than a single state duration. Therefore even if the prior is correctly assigned and payoff comparisons are perfectly informative about the state of the world, sampling from population play fails to keep track of state changes with myopic agents. In evolutionary game theory terms, strictly dominated strategies do not vanish in the stationary distribution implied by any payoff-monotonic regular dynamics.

If the state changes occur with different probabilities, or if the gains from choosing the dominant action do not coincide in the two states, the players will not always adopt imitative strategies. Suppose in fact that the state changes from  $\omega_l$  to  $\omega_i$  with rate  $\lambda_i$ , and, say, that  $\lambda_2 > \lambda_1$ . Then a player adopting the optimal decision rule may in some cases play action  $a_2$  even though at the previous match he or she played  $a_1$  and tied with the opponent. The main result of the section, Proposition 4.2, however continues to hold in the sense that the population play concentrates on the action  $a_2$  in the long run. Therefore the  $\omega_1$ -durations will be  $\epsilon$ -undetected for all  $\epsilon > 0$ . This holds also in the case that  $\lambda_1 = \lambda_2$ , but the gain for playing  $a_2$  on an  $\omega_2$ -duration is larger than the gain for playing  $a_1$  on an  $\omega_1$ -duration. In sum, dominated actions are adopted in the long run by a majority of the population of myopic players for fractions at least  $\min\{\lambda_1/(\lambda_1 + \lambda_2), \lambda_2/(\lambda_1 + \lambda_2)\}$  of the total time. Additional details are provided in Appendix B.

## 5. FORWARD-LOOKING OPTIMIZATION

In this section, we assume that each player cares about his or her current own payoff as well as the future payoffs. At the same time, for analytical tractability, we use stationary strategies with a single period memory.<sup>7</sup> At the end of this section, we show that such an assumption can be justified if it is costly to keep track of the history more accurately. As in the previous

<sup>7</sup> The main results of the paper, in particular the existence of stationary equilibrium and the limiting characterization as  $\lambda \rightarrow 0$ , are valid with  $k$ -period memories as well.

section, we present only the case for  $\lambda_1 = \lambda_2$ .<sup>8</sup> Unlike the previous section, we focus the analysis on the case where the population matches are very frequent relative to the state changes.

In this section, the history observable to  $j$  at  $t$  is

$$h^j(t) = (a^j(m^j(t)), a^{j(m^j(t))}(m^j(t)), u^j(m^j(t))).$$

In fact, some of the information is superfluous since the action of the opponent can be deduced from the payoff realization. Therefore it is more convenient to define the history as  $h^j(t) = (a^j(m^j(t)), u^j(m^j(t)))$ . Notice that we are assuming here that players do not know  $m^j(t)$ , i.e., the strategies do not depend on calendar time. Apart from the differences in  $h^j(t)$ , the pure and mixed strategies are defined as in the previous section.

In order to simplify the calculations, we use the overtaking criterion, rather than the discounted sum of payoffs, for evaluating sequences of outcomes.<sup>9</sup> Formally, let  $\{m_k\}_{k=0}^\infty \equiv \mathbf{m}$  be the random sequence of future matching times for  $j$ . The sequence of future actions chosen by  $j$  is then denoted by  $\{a^j(m_k)\}_{k=0}^\infty \equiv \mathbf{a}_j$ , and the actions taken by  $j$ 's opponents are denoted by  $\{a^{j(m_k)}(m_k)\}_{k=0}^\infty \equiv \mathbf{a}_{j(\mathbf{m})}$ . To evaluate the utilities from various action profiles, we consider the following infinite summations:

$$\pi(\mathbf{a}_j, \mathbf{a}_{j(\mathbf{m})}) = \sum_{k=0}^{\infty} u(a^j(m_k), a^{j(m_k)}(m_k), \omega(m_k)).$$

If the summation above does not converge, assign the value  $-\infty$  to  $\pi$ . Since the players are randomly matched, an expectation must be taken over the future opponents when evaluating the payoffs.

Notice that different initial choices of actions induce different distributions on continuation plays for each player. Let the future actions of player  $j$  conditional on an initial choice  $a_i$  be denoted by  $\mathbf{a}_j^i$  so that a choice at matching instant  $m_k$  following initial choice  $a_i$  is given by  $a_i^j(m_k)$ . Let  $\mathbf{m}$ ,  $\mathbf{a}_j^i$ , and the actions of future opponents,  $\mathbf{a}_{j(\mathbf{m})}$ , be drawn from their respective distributions.

According to the overtaking criterion, player  $j$  prefers action  $a_1$  to action  $a_2$  if there is a  $\bar{K} < \infty$  such that for all  $K \geq \bar{K}$ ,

$$E \sum_{k=0}^K u(a_1^j(m_k), a^{j(m_k)}(m_k), \omega(m_k)) \geq E \sum_{k=0}^K u(a_2^j(m_k), a^{j(m_k)}(m_k), \omega(m_k)),$$

<sup>8</sup> The case for  $\lambda_1 \neq \lambda_2$  yields similar results. The Appendix, when presenting the proofs of the statements presented in this section, also points out the differences for the case when  $\lambda_1 \neq \lambda_2$ .

<sup>9</sup> The limit of means criterion does not discriminate enough between sequences of outcomes for our purposes, because the effect of any individual decisions is vanishing in the limit (the processes that result from the analysis are strongly mixing).

where the expectations are taken with respect to the random matching probabilities. In the last part of this section, the impact of the current choice on the future choices and payoffs is made explicit. In solving for optimal strategies, we use the recursive formulation based on strong long run averages as defined in Dutta [11].<sup>10</sup> The definition of an equilibrium strategy profile  $\sigma$  is the same as in the previous section.

We solve for the symmetric stationary equilibrium strategies of the game. By using the notation  $\sigma$  to denote the mixed strategy of any given player, we are looking for a strategy  $\sigma$  such that it is optimal for each player to use  $\sigma$  if all the other players use  $\sigma$ . Notice that here we are assuming that a player's own future choices comply with  $\sigma$ .

The first two results of this section make the case for imitation and experimentation. The following lemma, in particular, shows that the optimal strategy must yield imitation after a history that reveals the state of the world. The notation  $\sigma_i(h^j(t))$  identifies the probability that any player  $j$  plays  $a_i$  after history  $h^j(t)$ .

**PROPOSITION 5.1.** *For any  $\lambda$  there is a  $\mu(\lambda)$  such that whenever  $\mu \geq \mu(\lambda)$ , in equilibrium,  $\sigma_i(a_i, 1) = 1$  and  $\sigma_i(a_i, -1) = 1$ , for any pair of actions  $(a_i, a_i)$  with  $a_i \neq a_i$ .*

While the imitation rule is settled as the optimal strategy when the history reveals the true state of the world, the next result establishes the value of experimentation after histories that do not reveal the state of the world. As long as the states do not change too often, there does not exist an equilibrium where players play imitation after any such histories.

**PROPOSITION 5.2.** *For any  $\lambda$  there is a  $\mu(\lambda)$  such that whenever  $\mu \geq \mu(\lambda)$ , in any stationary symmetric equilibrium,  $\sigma_i(a_i, 0) < 1$  for any action  $a_i$ .*

The intuition behind this result is quite simple. If all the other players are using the pure imitation rule, then it is optimal for a given individual to change his or her action conditional on observing a tie in his or her previous match. The reason for this is that by the results in the previous section, the play within the population does not react to most state changes in the long run. If the population is currently concentrated on the inferior action, a single trial leads to a large expected number of wins in the future. If the population is concentrated on the dominant action, a change of

<sup>10</sup> In what follows, we show that strategies that are optimal according to this criterion are optimal under overtaking criterion as well, since the process of payoff realizations under the optimal strategy is stationary and ergodic. This follows since the conditions for Theorem 5 in Dutta [11] are satisfied. This same theorem shows that the equilibrium policies can be regarded as limits of the equilibrium policies of a discounted model, as the discount factor tends to unity.

actions leads to a single period loss. Therefore as long as the matching rate  $\mu$  is high enough, the gains persist over many periods, while the losses take place in a single period, and it is optimal to experiment occasionally.

Given that the state changes with the same rate from  $\omega_1$  to  $\omega_2$  and from  $\omega_2$  to  $\omega_1$ , it is meaningful to restrict attention to equilibria where  $\sigma_1(a_1, 0) = \sigma_2(a_2, 0)$  and we introduce  $\varepsilon = 1 - \sigma_i(a_i, 0)$ . If the state is not revealed in the previous match, then the previous action is chosen again with probability  $1 - \varepsilon$ .

For each fixed  $\mu$ , and for any choice of the experimentation parameter  $\varepsilon$ , we can derive the law of motion for the population choices of actions. Let  $x_\varepsilon(t)$  denote the fraction of players choosing action  $a_1$  in period  $t$ . Note that we are parameterizing the process of population play by the relevant experimentation probabilities. In state  $\omega_1$ , we have:

$$\begin{aligned}\dot{x}_\varepsilon(t) &= \mu x_\varepsilon(t)(1 - x_\varepsilon(t)) + \mu\varepsilon(1 - x_\varepsilon(t))^2 - \mu\varepsilon x_\varepsilon(t)^2 \\ &= \mu[\varepsilon + x_\varepsilon(t)(1 - 2\varepsilon) - x_\varepsilon(t)^2].\end{aligned}\tag{4}$$

It is easy to calculate the long run level of  $x_\varepsilon(t)$  in the case where the state does not change. For this, we simply set the rate of change in the above equation equal to zero and solve for the stationary  $x_\varepsilon$ . The relevant root of the quadratic equation is:

$$\bar{x}_\varepsilon = \frac{1 - 2\varepsilon + \sqrt{1 + 4\varepsilon^2}}{2}.\tag{5}$$

The same reasoning leads to the following law of motion and the corresponding long run steady state in state  $\omega_2$

$$\dot{x}_\varepsilon(t) = \mu[\varepsilon - x_\varepsilon(t)(1 + 2\varepsilon) + x_\varepsilon(t)^2],\tag{6}$$

and

$$\underline{x}_\varepsilon = \frac{1 + 2\varepsilon - \sqrt{1 - 4\varepsilon^2}}{2}.\tag{7}$$

Notice that for  $\varepsilon > 0$ ,  $\bar{x}_\varepsilon < 1$ , and  $\underline{x}_\varepsilon > 0$ . In other words, the process of population play is bounded away from the boundary of the unit interval. This induces a qualitative change in the behavior of the system as compared to the case with pure strategies. For example, it is easy to see that  $x(t)$  has a unique invariant distribution on the open interval  $(\underline{x}_\varepsilon, \bar{x}_\varepsilon)$ .<sup>11</sup>

<sup>11</sup> Unfortunately the calculation of the invariant distribution is not an easy matter. For general results on stochastic processes of the type described above, see, e.g., Davis [8].

This is in sharp contrast with the pure strategy case where the process spends asymptotically all of its time arbitrarily close to the boundary of  $[0, 1]$ .

An easy intuition for the difference in the results is the following. By introducing the randomization, the time symmetry in the process is broken. In particular, in state  $\omega_1$ , the rate of increase of  $x(t)$  approaches 0 as  $x(t)$  converges to  $\bar{x}_\varepsilon$ . On the other hand, in state  $\omega_2$ , the rate of decrease of  $x(t)$  (i.e., also the rate of increase of action  $a_2$ ) at  $\bar{x}_\varepsilon$  is bounded away from zero for all  $\varepsilon > 0$ .<sup>12</sup>

In order to start the analysis of the individual decision problem, we need to make an assumption about the initial distribution of the action profile in the population as well as the initial state  $\omega(0)$ . Since we do not want to give any particular significance to the initial period and since the joint process  $(x(t), \omega(t))$  is ergodic on  $(\bar{x}_\varepsilon, \bar{x}_\varepsilon) \times \Omega$ , a natural initial condition seems to be that all variables are drawn from the relevant invariant distribution. The implicit assumption then is that this game has been played for an arbitrarily long history prior to the start of the analysis. A consequence of this modeling choice is that the decision problem of all the individuals is the same prior to observing the outcome in the previous match.

The key observation for the analysis of optimal individual decisions is that the process determining the future opponents of a player and the population shares of the actions in the population at the matching times are not influenced by the past actions of the player. The following lemma states formally that the difference in the distribution of the continuation play of an individual player induced by a different initial action choice vanishes in finite time.

**LEMMA 5.1.** *For almost all  $\mathbf{m}$ ,  $\mathbf{a}_j^i$ , and  $\mathbf{a}_{j(m)}$ , there exists a  $K < \infty$  such that  $a_1^i(m_k) = a_2^i(m_k)$ , for all  $k > K$ . Furthermore,  $EK < \infty$ , where the expectation is taken with respect to the distribution of  $\mathbf{m}$ ,  $\mathbf{a}_j^i$ , and  $\mathbf{a}_{j(m)}$ .*

Since the payoffs are evaluated according to the overtaking criterion, we can concentrate on the differences in the payoffs during the first  $K$  periods. We start by showing that the game has no symmetric pure strategy equilibria. Recall that any proposed symmetric stationary equilibrium profile is characterized by a single parameter,  $\varepsilon$ . The exogenous environment is parameterized by  $(\lambda, \mu)$ . We hold  $\lambda$  fixed throughout the discussion and let  $\mu$  vary. This is without loss of generality since any model with parameters  $(p\lambda, p\mu)$  is equivalent to  $(\lambda, \mu)$  apart from a linear scaling in the units of measurement for time. Fix an arbitrary player  $j$  and denote by  $\alpha^\mu(\varepsilon)$  his

<sup>12</sup> The exact laws of motion in the two states can be solved by performing a simple change of variables. Since the formulas are not used later, they are omitted here.

or her set of optimal experimentation probabilities when all others, as well as the player in the future periods, experiment at rate  $\varepsilon$ , and the rate of matches is  $\mu$ .

**LEMMA 5.2.** *There is a  $\bar{\mu}$  such that for all  $\mu \geq \bar{\mu}$ ,  $\alpha^\mu(0) = 1$ .*

As a result, we conclude that zero experimentation is not a symmetric equilibrium. The next lemma shows that the rate of experimentation in a symmetric equilibrium cannot be very high if the frequency of matches is high.

**LEMMA 5.3.** *For any  $\bar{\varepsilon} > 0$ , there is a  $\bar{\mu}(\bar{\varepsilon})$  such that  $\alpha^\mu(\varepsilon) = 0$  for all  $\varepsilon \geq \bar{\varepsilon}$  and  $\mu \geq \bar{\mu}$ .*

The intuition for this result is also quite straightforward. If there is sufficient heterogeneity in the population, it is very unlikely for a player to realize the benefits from an experiment for a long string of matches. At the same time, the action that resulted in a draw is more likely to be the dominant action, and since a (relatively) large fraction of the opponents are experimenting, the myopic gain from not experimenting is quite high.

The payoff function of player  $j$  is continuous in the population experimentation rate  $\varepsilon$  since it is a time integral of a payoff that is continuous in  $\varepsilon$  against a Poisson arrival process. Thus Lemma 5.2, Lemma 5.3, and a simple application of the intermediate value theorem allow us to conclude the main existence result of this section.

**PROPOSITION 5.3.** *For all  $\mu$ , there is an  $\varepsilon > 0$  such that  $\varepsilon \in \alpha^\mu(\varepsilon)$ . Moreover,  $\lim_{\mu \rightarrow \infty} \underline{\varepsilon}(\mu) = 0$ , where  $\underline{\varepsilon}(\mu) = \sup\{\varepsilon \mid \varepsilon \in \alpha^\mu(\varepsilon)\}$ .*

In words, we have demonstrated the existence of symmetric equilibria. Furthermore, we have shown that for large  $\mu$ , the equilibrium experimentation probabilities are small. The remainder of this section investigates the asymptotic rate at which  $\varepsilon$  converges to zero as  $\mu$  increases. This exercise is essential if we want to determine how well coordinated the population is on the dominant action in the long run, as the state changes become very rare in comparison to the matches.

In order to obtain estimates on the rate of convergence, it is useful to look at an auxiliary random process that approximates the population process  $x(t)$  for large  $\mu$ . The key to the approximation that we perform is the observation that the real time that it takes for the frequency of action  $a_1$  to grow from an arbitrarily low level  $\delta$  to  $1 - \delta$  is extremely short for large  $\mu$ . As a result, for large  $\mu$ ,  $x(t)$  spends most of its time close to 0 or 1.

Hence we approximate the process  $x(t)$  by a simpler process that lives on the two asymptotic values calculated above for the real population process.

Let  $\hat{x}^\mu(t) \in \{\underline{x}_\varepsilon, \bar{x}_\varepsilon\}$  be the approximate population process. To make the approximation valid as  $\mu \rightarrow \infty$ , we need to describe how much time is spent in each of the two possible states. Let  $T(\mu, \varepsilon)$  be the amount of real time that the approximating process spends in state  $\underline{x}_\varepsilon$ . The approximation is valid if we require that  $T(\mu, \varepsilon)$  equals the amount of time that it takes for the population to increase from  $\underline{x}_\varepsilon$  to  $1/2$ . At the same time, we must make sure that  $T(\mu, \varepsilon)$  is such that each player is indifferent between experimenting and not experimenting. Combining these two requirements, we obtain a characterization of the aggregate equilibrium behavior as  $\mu \rightarrow \infty$ .

**PROPOSITION 5.4.** *For any  $\varepsilon(\mu)$  such that  $\varepsilon(\mu) \in \alpha^\mu(\varepsilon(\mu))$ ,*

$$\lim_{\mu \rightarrow \infty} \mu T(\mu, \varepsilon(\mu)) = O(\sqrt{\mu}) \quad (8)$$

$$\lim_{\mu \rightarrow \infty} \varepsilon(\mu) \approx \frac{1}{2} e^{-\sqrt{2\mu}}. \quad (9)$$

The validity of the approximation used to get this result is also shown in the Appendix. The message of the theorem is clear. Since the total expected number of matches grows linearly in  $\mu$ , and since the number of matches before a  $\frac{1}{2}$ -detected (and hence also  $\gamma$ -detected for any  $\gamma < \bar{x}_\varepsilon$ ) state change grows linearly in  $\sqrt{\mu}$ , almost all the players are choosing the dominant action almost all of the time when  $\mu \rightarrow \infty$ . Thus we are close to full optimality in a qualitative sense even though the public goods nature of experimentation leads to some suboptimality.

In the remainder of the section, we sketch an argument to show that it is not in any player's interest to buy costly information about past matches. With that in mind, we may interpret the model as one of endogenously imperfect recall.

Fix as a baseline the full information optimal strategy: play  $a_i$  at  $t$  if and only if  $\omega(t) = \omega_i$ . Consider an  $\omega_1$ -duration  $(\tau_k, \tau_{k+1})$ . Set  $\lambda = 1$ , and for any small  $\delta > 0$ , let  $T$  be such that  $x(T + \tau_k) = \bar{x}_\varepsilon - \delta$ . Against the bounded-memory equilibrium population dynamics, the optimal-strategy average payoff on the state duration is bounded above by  $[T + (1 - \bar{x}_\varepsilon + \delta)(1 - T)]$ . For  $\mu$  large enough, we know that  $T \approx 0$ , and that  $\bar{x}_\varepsilon \approx 1 - \varepsilon \rightarrow 0$ , so that the payoff from using the optimal strategy is bounded above by a value close to  $\delta$ . If the players were able to purchase full information in each period at cost  $C$ , their optimal average payoff would thus be bounded above by a value close to  $\delta - C$ . For  $\mu$  large enough,  $x(\tau_{k+1}) \approx \bar{x}_\varepsilon$  almost surely, so that we can choose  $\delta \approx 0$ .

Consider the average payoff when using the bounded-memory equilibrium strategy. By revealed preference we know that such a payoff is not smaller than the average payoff obtained by a player using the pure imitation rule. This payoff is bounded below by  $-1/\mu$ ; as with probability close to 1, player  $j$  will face an opponent taking  $a_1$ , receive a payoff of  $-1$ , and play  $a_2$  thereafter. For an arbitrary  $C > 0$ , we can choose  $\mu$  large enough to have  $-\frac{1}{\mu} > -C$ .

## 6. CONCLUSION

In this paper, we considered the evolution of play in a changing environment. The particular model was chosen to reflect the idea that players can learn from relative payoff comparisons, but not from their absolute stage game payoffs. A more realistic assumption would be to allow for some learning from one's own past choices regardless of the actions that other agents chose. The techniques developed here would be useful for those models too, as long as social learning is not swamped by learning from one's own past experiences.

Consider, for example, a model where the players choose between a safe action whose payoff is independent of the state of the world and an uncertain action that yields a high payoff in one of the states and a low payoff in the other. Assume also that prior to choosing their next action, the players observe the action and the payoff of a randomly selected player in the population. Using the techniques of this model, we could show that equilibria of that model are approximately efficient as the state changes become infrequent enough.

## APPENDIX A

### *Proofs Omitted from Section 4*

*Proof of Proposition 4.1.* We show that regardless of the opponents' strategies, the strictly dominant strategy of each player  $j$  is to adopt the pure imitation rule. It follows that in the unique equilibrium all players adopt the pure imitation rule.

Pick any player  $j$  and time  $t$ . Set

$$m_*^j(t) = \sup\{m < t : u_j(a^j(m^j(t)), a^{j(m)}(m^j(t))) \neq 0\},$$



and

$$\underline{k} = \sup\{k: \tau_k < m_*^j(t)\}, \bar{k} = \sup\{k: \tau_k < t\}, \text{ and } K = \bar{k} - \underline{k}.$$

Setting  $\omega_i = \omega(m_*^j(t))$ , the optimal action of player  $j$  at  $t$  coincides with  $a_i$  if  $K$  is an even number, and it coincides with the opposite action if  $K$  is an odd number. Since the state changes are governed by a Poisson process, for any  $x \leq t$ , the renewal system of  $K$  is

$$\Pr(0; x, t) = e^{-\lambda(t-x)}$$

and

$$\Pr(K; x, t) = \int_x^t \lambda e^{-\lambda s} \Pr(K-1; s, t) ds$$

iteratively for any  $K > 0$ . Recursive calculations show that for any  $t$ , any  $m_*^j(t)$ , and any  $l \geq 0$ , it must be the case that  $\Pr(2l; m_*^j(t), t) > \Pr(2l+1; m_*^j(t), t)$ . Therefore player  $j$  plays  $a_i$  at time  $t$ . This means that  $j$  adopts the imitation rule at  $t$ .

*Proof of Lemma 4.1.* Given the monotonicity of  $x_A$  in the set inclusion, it is sufficient to check that the claim holds for all closed intervals,  $A = [\epsilon, 1 - \epsilon]$ .

Given the random sequence of state change times  $\{\tau_i\}_{i=0}^\infty$ , in order to derive the asymptotic properties of the process  $x$ , it is sufficient to calculate  $x(t)$  on  $\{\tau_i\}_{i=0}^\infty$ . Start the system without loss of generality at  $x(0) = 1/2$ ,  $\omega(0) = \omega_1$ . For  $i$  odd, the switches are from  $\omega_1$  to  $\omega_2$  and for  $i$  even, they are from  $\omega_2$  to  $\omega_1$ . It follows  $x(\tau_i) < x(\tau_{i+1})$  for  $i$  even and the reverse inequality holds for  $i$  odd. Define the sequence of random durations  $\{\sigma_k\}_{k \geq 1}$ ,  $\{y_s\}_{s \geq 1}$ , and  $\{\beta_k\}_{k \geq 1}$  as follows. For any  $s \geq 1$ , set  $\Delta_{1s} = \tau_{2s-1} - \tau_{2(s-1)}$  and  $\Delta_{2s} = \tau_{2s} - \tau_{2s-1}$ . As the state changes are characterized by a Poisson arrival process of parameter  $\lambda$ , it follows that for  $i = 1, 2$ ,  $\Delta_{is} \sim \exp(1/\lambda)$ , i.i.d. For any  $s \geq 1$ , and  $k \geq 1$ , put

$$y_s = \Delta_{1s} - \Delta_{2s}, \quad \sigma_k = \sum_{s=1}^k y_s, \quad \beta_k = \sigma_k + \Delta_{2k}.$$

The sequence  $\{\sigma_k\}$  is a driftless random walk with a strictly positive, constant, bounded variance. Notice that for any  $k$ , we can express  $\beta_k$  as  $\Delta_{10} + \gamma_k$ , where  $\{\gamma_k\}$  is also a driftless random walk with a strictly positive, constant, bounded variance. By construction,

$$x(\tau_{2k}) = x(\sigma_k) = \frac{1}{1 + \frac{1-x_0}{x_0} e^{-\mu\sigma_k}} \tag{A1}$$

$$x(\tau_{2k-1}) = x(\beta_k) = \frac{1}{1 + \frac{1-x_0}{x_0} e^{-\mu\beta_k}}.$$

For any  $\epsilon$ , choose  $K(\epsilon)$  such that

$$x(K(\epsilon)) = \frac{1}{1 + \frac{1-x_0}{x_0} e^{-\mu K(\epsilon)}} \geq 1 - \epsilon.$$

This implies that  $x(\tau_{2k-1}) > 1 - \epsilon$  if  $\beta_k > K(\epsilon)$ , and a fortiori if  $\gamma_k > K(\epsilon)$ . By symmetry, this also implies that  $x(\tau_{2k}) < \epsilon$  whenever  $\sigma_k < -K(\epsilon)$ . By the contrapositive of the Wald equation (see [10], the expected time of re-entry of  $\sigma_k$  from  $(K(\epsilon), \infty)$  to  $(-\infty, K(\epsilon)]$  is infinite. Conversely, also the expected time of re-entry of  $\gamma_k$  from  $(-\infty, -K(\epsilon))$  to  $[-K(\epsilon), \infty)$  is infinite.

Starting inside  $[-K(\epsilon), K(\epsilon)]$ , the expected hitting time for  $\sigma_k$  (and also for  $\gamma_k$ ) to  $(K(\epsilon), \infty) \cup (-\infty, -K(\epsilon))$  is bounded above by  $K(\epsilon)^2/E[y_1^2] < \infty$ . Since  $\sigma_k^2 - kE[y_1^2]$  is a martingale, in fact, it follows that for any  $k$ ,  $E[\sigma_k^2 - kE[y_1^2]] = 0$ . Setting  $T = \inf\{k: \sigma_k \notin [-K(\epsilon), K(\epsilon)]\}$ , this implies that  $K(\epsilon)^2 = E[\sigma_T^2] = E[T] E[y_1^2]$ .

We have shown that

$$\lim_{T \rightarrow \infty} \frac{\int_0^T I_{[\epsilon, 1-\epsilon]}(x(t)) dt}{T} = 0 \text{ a.s.}$$

*Proof of Proposition 4.2.* We need to show that for all  $\epsilon > 0$ ,

$$\lim_{l \rightarrow \infty} \frac{\#\{l: x(t) \leq \epsilon \text{ for } \tau_{2l} \leq t \leq \tau_{2l+1}\} + \#\{l: x(t) \geq 1 - \epsilon \text{ for } \tau_{2l+1} \leq t \leq \tau_{2(l+1)}\}}{2l}$$

$$= \frac{1}{2}.$$

Choose  $K(\epsilon)$  to be such that  $\beta_k > K(\epsilon)$ ; it follows that  $x(\beta_k) > 1 - \epsilon$ . By Lemma 4.1, we know that  $x(t)$  spends all of its total time in  $(0, \epsilon) \cup (1 - \epsilon, 1)$ . The claim is then true unless the process crosses from  $(0, \epsilon)$  to  $(1 - \epsilon, 1)$  on a positive fraction of the total state changes. But this would contradict  $x_{[\epsilon, 1-\epsilon]} = 0$ .

## APPENDIX B

*Myopic Players with  $\lambda_1 \neq \lambda_2$* 

In this section we characterize the optimal strategy of myopic players when the state changes occur with different probabilities and show that the population will concentrate on the action that corresponds to the state with highest flow rate, so that all durations in the state with the smallest flow rate will be undetected.

The determination of the optimal decision of players conditioning their choice on the complete history of play and on calendar time when  $\lambda_1 \neq \lambda_2$  is rather involved and is thus postponed to further research. For our current purposes, it suffices to point out that it will not consist of purely imitative strategies.

First note that if  $\lambda_1$  is very dissimilar to  $\lambda_2$ , the optimal strategy cannot be imitation. It is in fact straightforward to notice that as  $\lambda_2/\lambda_1 \rightarrow \infty$ , and  $\mu$  is fixed, the problem approximates one where no significant uncertainty about the environment occurs, and so the optimal strategy is to play always  $a_2$ , the dominant strategy when the true state of the world is  $\omega_2$ . Also, since strategies depend on calendar time, for any triple  $(\lambda_1, \lambda_2, \mu)$ , one can find histories after which players will not adopt the imitation. Suppose in fact that  $\lambda_2 > \lambda_1$ ; pick any player  $j$  and time  $t$  such that  $a^j(m^j(t)) = a_1$  and  $u^j(m^j(t)) = 0$ . For  $m^j(t)$  large enough, since the process is strongly mixing, the probability that  $\omega(t) = \omega_2$  approximates  $\lambda_2/[\lambda_1 + \lambda_2]$ , and thus player  $j$  will play action  $a_2$ . On the other hand, a straightforward extension of the proof of Proposition 4.1 yields that in equilibrium  $\sigma_2^j(h^j(t)) = 1$  as long as  $h^j(t) \notin L_2^j(t)$ . Player  $j$  never abandons action  $a_2$  unless it was defeated at the previous match.

The next proposition shows that if  $\lambda_2 > \lambda_1$ , then the  $\omega_1$ -durations are not detected by the population dynamics induced by pure imitation. This result holds a fortiori for the equilibrium population dynamics, because we have shown above that in equilibrium the players play  $a_1$  on a set of histories which is not larger than the set of histories where  $a_1$  is prescribed by the pure imitation rule.

**PROPOSITION B.1.** *For any  $\mu$ , when all players adopt the pure imitation rule, if  $\lambda_2 > \lambda_1$  then  $\lim_{t \rightarrow \infty} x(t) = 0$  a.s., and if  $\lambda_1 > \lambda_2$  then  $\lim_{t \rightarrow \infty} x(t) = 1$  a.s.*

*Proof.* Recall the construction of the random durations  $A_{is}$ ,  $\{\sigma_k\}$ , and  $\{y_s\}$ , from the proof of Lemma 4.1. Since for any  $i = 1, 2$ , and  $l \neq i$ ,  $A_{is} \sim \exp(1/\lambda_l)$ , i.i.d, the sequence  $\{\sigma_k\}$  is a random walk with drift and strictly positive, constant, bounded variance. The recurrence properties of

this walk depend on the sign of the mean of  $y_s$ . If this mean is strictly negative, i.e.,  $\lambda_2 > \lambda_1$ , then by the strong law of large numbers, for all  $K$ ,  $\Pr\{\sigma_k > K \text{ for infinitely many } k\} = 0$ . By construction,

$$x(\tau_{2k}) = x(\sigma_k) = \frac{1}{1 + \frac{1-x_0}{x_0} e^{-\mu\sigma_k}}.$$

For any  $\varepsilon$ , choose  $K(\varepsilon)$  such that

$$x(K(\varepsilon)) = \frac{1}{1 + \frac{1-x_0}{x_0} e^{-\mu K(\varepsilon)}} \leq 1 - \varepsilon.$$

Since  $\sigma_k < K(\varepsilon)$  for all but finitely many  $k$ , the almost sure convergence of the process  $x(t)$  to 0 follows. A similar construction applies to the case where  $\lambda_1 > \lambda_2$ . ■

## APPENDIX C

### *Proof Omitted from Section 5*

*Proof of Proposition 5.1.* Consider any arbitrary player  $j$  at any time  $t$ , such that  $u_j(a^j(m^j(t)), a^{j(m)}(m^j(t))) \neq 0$ . Setting  $\tau' = \inf\{\tau_k > m^j(t)\}$ , and  $\omega_i = \omega(m^j(t))$ , the optimal action of player  $j$  at  $t$  coincides with  $a_i$  if  $t < \tau'$ . Since matches and state changes are governed by a Poisson process,

$$\Pr\{t < \tau'\} = \int_0^\infty \left[ \int_s^\infty \lambda_l e^{-\lambda t} dt \right] \mu e^{-\mu s} ds = \frac{\mu}{\mu + \lambda_l} \rightarrow 1, \quad \text{for } \mu \rightarrow \infty,$$

where  $\lambda_l$  is the flow rate from state  $\omega_i$  to state  $\omega_l$ ,  $l \neq i$ .

*Proof of Proposition 5.2.* Consider first the case for  $\lambda_1 \neq \lambda_2$ , and say without loss of generality that  $\lambda_2 > \lambda_1$ . Proceed by contradiction, and suppose that the symmetric adoption of the pure imitation rule is a stationary equilibrium. Let  $x$  denote the induced population dynamics. In the proof of Proposition 2.1, we have shown that, for all  $\varepsilon > 0$ , there is a finite time  $T(\varepsilon)$  such that  $x(t) \in (0, \varepsilon)$  for any  $t \geq T(\varepsilon)$ .

Consider a triple of consecutive state changes  $\{\tau_{i-1}, \tau_i, \tau_{i+1}\}$  such that  $\tau_{i-1} > T(\varepsilon)$  and  $\omega(t) = \omega_1$  for any  $t \in (\tau_{i-1}, \tau_i)$ . We want to determine  $\Delta u(a_2, 0)$ , the expected difference in utility for playing  $a_1$  instead of  $a_2$ , on histories where the pure imitation rule prescribes  $a_2$ .

Say that a player  $j$  is matched at a time  $t \in (\tau_{i-1}, \tau_{i+1})$  such that  $h^j(t) = (a_2, 0)$ . Let  $\Delta u(a_2, 0 | \omega(t), t)$  be the expected difference in utility for playing  $a_1$  instead of  $a_2$  at  $t$ . First say that  $t \in (\tau_{i-1}, \tau_i)$ . Consider the random sequence of future matches  $\{m_k\}_{k \geq 1}$ , and let  $\bar{m} = \inf\{m_k : m_k > \tau_i\}$ . Since  $\tau_{i-1} > T(\epsilon)$ , with probability close to 1, by playing  $a_2$  instead of  $a_1$ , player  $j$  gains a payoff of 1 at  $t$  and plays  $a_1$  at match  $m_1$ . Proceeding recursively, with probability close to 1,  $j$  plays  $a_1$  at any  $m_k \leq \bar{m}$  and plays  $a_2$  at any  $m_k > \bar{m}$ . For  $\mu$  large enough, it follows that  $\Delta u(a_2, 0 | \omega_1, t) \approx \mu[\tau_i - t]$ .

Second say that  $t \in (\tau_i, \tau_{i+1})$ , by playing  $a_1$  instead of  $a_2$ , player  $j$  loses a payoff of 1 with probability close to 1, and he or she reverts to  $a_2$  at  $m_1$ . It follows that  $\Delta u(a_2, 0 | \omega_2, t) \approx -1$ .

Solving a simple compound renewal (Poisson) process, we obtain that  $\Pr(t \in (\tau_i, \tau_{i+1})) = \lambda_1 / [\lambda_1 + \lambda_2]$  and  $\Pr(t \in (\tau_{i-1}, \tau_i)) = \lambda_2 / [\lambda_1 + \lambda_2]$ . Since matches and state changes are independent processes, we have for  $\mu$  large enough,

$$\Delta u(a_2, 0) \approx \left[ \frac{\lambda_2}{\lambda_1 + \lambda_2} \int_{\tau_{i-1}}^{\tau_i} \frac{\mu[\tau_i - t]}{\tau_i - \tau_{i-1}} dt - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right] > 0.$$

For  $\lambda_1 = \lambda_2$ , it is the case that  $\lambda_1 / [\lambda_1 + \lambda_2] = 1/2 = \lambda_2 / [\lambda_1 + \lambda_2]$ , that  $x(t) \in \{(0, \epsilon), (1 - \epsilon, 1)\}$  for almost all  $t$ , and that 1/2 of the state durations are  $\epsilon$ -undetected, for any  $\epsilon > 0$ . The argument presented for the case of  $\lambda_2 > \lambda_1$  can be repeated with no modifications if the duration  $(\tau_i, \tau_{i+1})$  is  $\epsilon$ -undetected. If it is  $\epsilon$ -detected, then  $\Delta u(a_2, 0 | \omega_1, t)$  is bounded below by  $-1$ . Compounding  $\epsilon$ -detected with  $\epsilon$ -undetected durations, we obtain that  $\Delta u(a_2, 0) \geq \mu^2/16 - 3/4 > 0$ , for  $\mu$  large enough.

*Proof of Lemma 5.1.* Consider the choice of an arbitrary player  $j$  at an arbitrary matching time  $t$  such that either  $h^j(t) = (a_1, 0)$  or  $h^j(t) = (a_2, 0)$ . First take the case when  $\omega(t) = \omega_1$  and  $h^j(t) = (a_2, 0)$ . For any future matching time  $m_k$ , we denote by  $x_k$  the population play  $x(m_k)$ ; note that  $x_k$  is independent of  $a^j(t)$ . Consider the sequence  $\{\tau'_k\}_{k=1}^\infty$  of state changes taking place after  $t$ , constructed in the proof of Proposition 5.1. For any  $k$ , the realizations of  $m_k$  and  $\tau'_k$  are independent of  $j$ 's choice at  $t$ .

The transition of  $j$ 's histories on the sequence  $\{m_k\}_{k=1}^\infty$  can be expressed as a Markov process  $\{X_k\}$  with initial state  $(a^j(t), u^j(t))$ . As long as  $m_k < \tau'_k$ ,  $\{X_k\}$  can be described as a Markov chain with states  $S = \{(a_1, 0), (a_2, 0), (a_1, 1), (a_2, -1)\}$ . We introduce the state space  $S^2$ , where for each  $(s_1, s_2) \in S^2$ ,  $s_1$  identifies a state on the continuation induced by  $a^j(t) = a_1$  and  $s_2$  a state on the continuation of  $a^j(t) = a_2$ . We denote by  $\{Y_k\}$  the joint process with state space  $S^2$ , generated by running two copies of  $\{X_k\}$  with different initial states  $(a^j(t), u^j(t))$ . Since

$h^j(t) = (a_2, 0)$ , the initial distribution of state  $Y_0$  assigns probability  $1 - x(t)$  to state  $(a_1, 1; a_2, 0)$  and probability  $x(t)$  to state  $(a_1, 0; a_2, -1)$ .

Define as *coupling event* the set of states  $(s_1, s_2)$  such that  $s_1$  and  $s_2$  induce the same equilibrium choice. Formally, let  $C = \{(s_1, s_2) : \sigma_1(s_1) = \sigma_1(s_2)\}$ . Conditional on the event  $C$ , the distribution on the continuation induced by  $a^j(t) = a_1$  is identical to the distribution on the continuation induced by  $a^j(t) = a_2$ . The process  $\{Y_k\}$  is thus summarized by a five-state nonautonomous Markov chain with state space  $S' = \{(a_1, 1; a_2, 0), (a_1, 0; a_2, -1), (a_2, -1; a_1, 0), (a_2, 0; a_1, 1), C\}$  and transition matrix

$$P = \begin{bmatrix} (1-x_k)(1-\varepsilon) & 0 & \varepsilon(1-x_k) & 0 & 0 \\ (1-\varepsilon)x_k & 0 & \varepsilon x_k & 0 & 0 \\ 0 & \varepsilon x_k & 0 & (1-\varepsilon)x_k & 0 \\ 0 & \varepsilon(1-x_k) & 0 & (1-\varepsilon)(1-x_k) & 0 \\ \varepsilon & 1-\varepsilon & 1-\varepsilon & \varepsilon & 1 \end{bmatrix}.$$

By inspecting the transition matrix  $P$ , we see that for any  $\varepsilon > 0$ , since  $0 < x_k < 1$  for any  $k$ , the process  $Y_k$  is nondecomposable, and  $C$  is the unique absorbing state. Setting  $K_1 = \inf\{k : Y_k = C\}$ , we obtain that for any initial state  $Y_0$ ,  $\Pr(K_1 < \infty | Y_0) = 1$  and that  $E(K_1) < \infty$ .

The process  $\{Y_k\}$  may also be used to describe the continuations of a player  $j$  with  $h^j(t) = (a_1, 0)$ , if for any  $(s_1, s_2) \in S$ ,  $s_1$  identifies a state on the continuation induced by  $a^j(t) = a_2$ , and  $s_2$  a state on the continuation of  $a^j(t) = a_1$ . The initial distribution of state  $Y_0$  would assign probability  $1 - x(t)$  to state  $(a_2, 0; a_1, 1)$  and probability  $x(t)$  to state  $(a_2, -1; a_1, 0)$ .

The joint process for the case that  $\omega(t) = \omega_2$  is analogously constructed and displays the same recurrence properties. These results imply that for almost all  $\mathbf{m}$ ,  $\mathbf{a}_j^i$ , and  $\mathbf{a}_{j(m)}$ , there exists a  $K < \infty$  such that  $a_1^i(m_k) = a_2^j(m_k)$ , for all  $k > K$ , and moreover  $EK < \infty$ .

*Proof of Lemma 5.2.* Already shown in the proof of Proposition 5.2.

*Proof of Lemma 5.3.* Consider the choice of an arbitrary player  $j$  at an arbitrary  $t$  such that either  $h^j(t) = (a_2, 0)$  or  $h^j(t) = (a_1, 0)$ . Given the random sequence  $\{x_k\}$  defined in the proof of Lemma 5.1, and the random time  $\tau'_1$  constructed in the proof of Proposition 5.1, set  $x_K = \sup\{x_k : x_k < \tau'_1\}$ , set  $x_0 = x(t)$ , and  $\mathbf{x} = \{x_0, x_1, \dots, x_K\}$ . Letting  $a_i = a^j(m^j(t))$ , we denote by  $\Delta u(h^j(t) | \omega(t), \mathbf{x}, \varepsilon)$  the expected difference of utility on the finite sequence  $\{t, m_1, \dots, m_K\}$  for playing action  $a_l$  at  $t$  instead of  $a_i$ , where  $l \neq i$ .

*Step 1.* For any  $\varepsilon \geq \bar{\varepsilon}$ ,  $\Delta U(a_2, 0 | \omega_1, \mathbf{x}, \varepsilon)$  admits a finite upper bound uniform in  $x$  and a strictly positive lower bound.

Consider the joint process  $\{Y_k\}$  derived in the proof of Proposition 5.1. We introduce  $\Delta U(a_1, 1; a_2, 0 | \omega_1, \mathbf{x}, \varepsilon)$  and  $\Delta U(a_1, 0; a_2, -1 | \omega_1, \mathbf{x}, \varepsilon)$ , respectively, as the expected difference of utility on the finite sequence  $\{t, m_1, \dots, m_K\}$  between playing  $a_1$  and  $a_2$  at  $t$  given that  $j$ 's opponent plays  $a_2$  and given that he or she plays  $a_1$ . Standard calculations that can be obtained upon request from the authors show that

$$\Delta U(a_1, 1; a_2, 0 | \omega_1, \mathbf{x}, \varepsilon) < \sum_{k=0}^M (1-\varepsilon)^k < \frac{1}{\varepsilon}$$

and

$$\Delta U(a_1, 0; a_2, -1 | \omega_1, \mathbf{x}, \varepsilon) < 1 + \varepsilon^2 \sum_{k=0}^{M-1} (1-\varepsilon)^k < 1 + \varepsilon.$$

Compounding the two cases with the initial distribution of  $Y_0$  we obtain

$$\Delta U(a_2, 0 | \omega_1, \mathbf{x}, \varepsilon) < x_0 \frac{1}{\varepsilon} + (1-x_0)(1+\varepsilon) < \frac{1}{\varepsilon} + 2,$$

so that there is a finite bound  $\bar{\Delta}$ , uniform in  $\mathbf{x}$ , such that  $\Delta U(a_2, 0 | \omega_1, \mathbf{x}, \varepsilon) < \bar{\Delta}$ .

With analogous calculations, we conclude that

$$\Delta U(a_2, -1; a_1, 0 | \omega_1, \mathbf{x}, \varepsilon) < (1-x_1)(\varepsilon^2 + \varepsilon - 1),$$

and

$$\Delta U(a_2, 0; a_1, 1 | \omega_1, \mathbf{x}, \varepsilon) < (1-2\varepsilon + \varepsilon^3)(x_1 - 1) - 1,$$

so that there is bound  $\hat{\Delta}(\mathbf{x}) < 0$  such that  $\Delta U(a_1, 0 | \omega_1, \mathbf{x}, \varepsilon) < \hat{\Delta}(\mathbf{x})$ .

For any  $\mathbf{x}$ , denote by  $1-\mathbf{x}$  the sequence  $\{1-x_k\}_{k=0}^K$  where  $x_k \in \mathbf{x}$ . By symmetry of the population strategies, it must be the case that  $\Delta U(a_2, 0 | \omega_1, \mathbf{x}) = -\Delta U(a_1, 0 | \omega_1, 1-\mathbf{x})$ , and thus that  $\Delta U(a_2, 0 | \omega_1, \mathbf{x}) > \hat{\Delta}(1-\mathbf{x}) > 0$ .

*Step 2.* Calculation of  $\Delta U(a_2, 0 | \varepsilon)$ .

By symmetry of the population strategies, it must be the case that  $\Delta U(a_2, 0 | \omega_2, 1-\mathbf{x}) = -\Delta U(a_2, 0 | \omega_1, \mathbf{x})$ . Thus, for any  $\delta \in (\underline{x}_\varepsilon, \bar{x}_\varepsilon)$ , we can expand  $\Delta U(a_2, 0 | \varepsilon)$  as follows (for notational ease, we shall drop  $\varepsilon$  from the formula).

$\Delta U(a_2, 0)$ 

$$\begin{aligned}
&= \Delta U(a_2, 0 \mid \omega_1) \Pr(\omega_1 \mid a_2, 0) + \Delta U(a_2, 0 \mid \omega_2) \Pr(\omega_2 \mid a_2, 0) \\
&= \int_{\{\mathbf{x} \mid x_0 < \bar{x}_\varepsilon - \delta\}} [\Delta U(a_2, 0 \mid \omega_1, \mathbf{x}) \Pr(\omega_1 \mid (a_2, 0), \mathbf{x}) \\
&\quad + \Delta U(a_2, 0 \mid \omega_2, \mathbf{x}) \Pr(\omega_2 \mid (a_2, 0), \mathbf{x})] d \Pr(\mathbf{x} \mid x_0 < \bar{x}_\varepsilon - \delta) \Pr(x_0 < \bar{x}_\varepsilon - \delta) \\
&\quad + \int_{\{\mathbf{x} \mid x_0 > \bar{x}_\varepsilon - \delta\}} [\Delta U(a_2, 0 \mid \omega_1, \mathbf{x}) \Pr(\omega_1 \mid (a_2, 0), \mathbf{x}) \\
&\quad + \Delta U(a_2, 0 \mid \omega_2, \mathbf{x}) \Pr(\omega_2 \mid (a_2, 0), \mathbf{x})] d \Pr(\mathbf{x} \mid x_0 > \bar{x}_\varepsilon - \delta) \Pr(x_0 > \bar{x}_\varepsilon - \delta) \\
&= \int_{\{\mathbf{x} \mid x_0 < \bar{x}_\varepsilon - \delta\}} \{ \Delta U(a_2, 0 \mid \omega_1, \mathbf{x}) [\Pr(\omega_1 \mid (a_2, 0), \mathbf{x}) - \Pr(\omega_2 \mid (a_2, 0), 1 - \mathbf{x})] \} \\
&\quad \times d \Pr(\mathbf{x} \mid x_0 < \bar{x}_\varepsilon - \delta) \cdot \Pr(x_0 < \bar{x}_\varepsilon - \delta) \\
&\quad + \int_{\{\mathbf{x} \mid x_0 > \bar{x}_\varepsilon - \delta\}} \{ \Delta U(a_2, 0 \mid \omega_1, \mathbf{x}) [\Pr(\omega_1 \mid (a_2, 0), \mathbf{x}) - \Pr(\omega_2 \mid (a_2, 0), 1 - \mathbf{x})] \} \\
&\quad \times d \Pr(\mathbf{x} \mid x_0 > \bar{x}_\varepsilon - \delta) \cdot \Pr(x_0 > \bar{x}_\varepsilon - \delta) \tag{C.1}
\end{aligned}$$

*Step 3.* For any  $\varepsilon \geq \bar{\varepsilon}$ , and any  $\delta \in (\underline{x}_\varepsilon, \bar{x}_\varepsilon)$ , there is  $\bar{\mu}(\bar{\varepsilon}, \delta)$  such that for all  $\mu \geq \bar{\mu}$ , the first term in Eq. (C.1) above vanishes.

Let  $x(t')$  denote the solution of the Cauchy problem identified by Eq. (4) and  $x(0) = \underline{x}_\varepsilon$ . Let  $T(\delta, \varepsilon, \mu)$  be the time  $t'$  such that  $x(t') = \bar{x}_\varepsilon - \delta$ . We know that  $T(\delta, \varepsilon, \mu) \rightarrow 0$  for  $\mu \rightarrow \infty$ . Set  $\tau_{\bar{i}} = \sup\{\tau_i < t\}$ , and set  $\hat{T}(\delta, \varepsilon, \mu) = \inf\{t' > \tau_{\bar{i}} : x(t') = \bar{x}_\varepsilon - \delta\} - \tau_{\bar{i}}$ .

For  $\mu$  large enough  $\hat{T}(\delta, \varepsilon, \mu)$  exists,  $\hat{T}(\delta, \varepsilon, \mu) < \tau_{\bar{i}+1} - \tau_{\bar{i}}$ , and  $\hat{T}(\delta, \varepsilon, \mu) < T(\delta, \varepsilon, \mu)$ . Since  $E(\tau_{\bar{i}+1} - \tau_{\bar{i}}) = 1/\lambda$ , it follows that  $\Pr(x_0 < \bar{x}_\varepsilon - \delta) \approx 0$  for  $\mu$  large enough. By Step 1,  $\Delta U(a_2, 0 \mid \omega_1, \mathbf{x}, \varepsilon)$  is bounded uniformly in  $\mathbf{x}$ , thus the first term in Eq. (C.1) vanishes.

*Step 4.* For any  $\varepsilon \geq \bar{\varepsilon}$ , the second term in Eq. (C.1) is strictly positive.

Consider the likelihood ratio:

$$\Lambda = \frac{\Pr(\omega_1 \mid (a_2, 0), \mathbf{x})}{\Pr(\omega_2 \mid (a_2, 0), 1 - \mathbf{x})} = \frac{\Pr(\omega_1, (a_2, 0), \mathbf{x})}{\Pr(\omega_2, (a_2, 0), 1 - \mathbf{x})}.$$

The symmetry of population strategies implies that  $\Pr(a_2, 0 \mid 1 - \mathbf{x}, \omega_2) = \Pr(a_1, 0 \mid 1 - \mathbf{x}, \omega_1)$  and that, since the match process and the state-change



process run independently,  $\Pr(1 - \mathbf{x} \mid \omega_2) = \Pr(\mathbf{x} \mid \omega_1)$ . The symmetry of the state-change process yields  $\Pr(\omega_2) = \Pr(\omega_1)$ . Thus, the relation

$$\Pr(\omega_2, (a_2, 0), 1 - \mathbf{x}) = \Pr(\omega_1, (a_1, 0), \mathbf{x})$$

must hold. Since both probabilities are bounded away from zero, we obtain that

$$A = \frac{\Pr(\omega_1, (a_2, 0), \mathbf{x})}{\Pr(\omega_1, (a_1, 0), 1 - \mathbf{x})} = \frac{\Pr(a_2, 0 \mid \omega_1, \mathbf{x})}{\Pr(a_1, 0 \mid \omega_1, 1 - \mathbf{x})} = \frac{1 - x_0}{x_0} < 1 - b$$

for some positive bound  $b$ . By Step 1,  $\Delta U(a_2, 0 \mid \omega_1, \mathbf{x}) > \hat{A}(1 - \mathbf{x}) > 0$ , it follows that  $\Delta U(a_2, 0, \varepsilon) < 0$ . Note that, by symmetry,  $\Delta U(a_1, 0, \varepsilon) = \Delta U(a_2, 0, \varepsilon) < 0$ . Recall the construction of the joint process  $\{Y_k\}$  in the proof of Proposition 5.2. We have shown that letting  $C$  be the coupling event, and setting  $K_1 = \inf\{k: Y_k = C\}$ , it is the case that  $\Pr(K_1 < \infty) = 1$ . For  $\mu$  large enough, it must be the case that  $\Pr(K_1 > K) \approx 0$ . Therefore  $\Delta U(a_2, 0, \varepsilon) < 0$  implies that for all  $K' > K_1$ ,

$$E \sum_{k=0}^{K'} u(a_2^{jk}, a^{jk(m_k)}, \omega(m_k)) > E \sum_{k=0}^{K'} u(a_1^{jk}, a^{jk(m_k)}, \omega(m_k)).$$

The overtaking criterion thus implies that  $\alpha(\varepsilon) = 0$ .

For  $\lambda_1 \neq \lambda_2$ , the above arguments can be extended as long as  $|\lambda_1 - \lambda_2| < M(\mu)$ , for some  $M(\mu)$ , where as in Lemma 5.1, the function  $M$  is strictly positive, strictly increasing, and  $M(\mu) \rightarrow \infty$  for  $\mu \rightarrow \infty$ .

*Proof of Proposition 5.3.* Consider the choice of a player  $j$  at an arbitrary time  $t$  such that  $h^j(t) = (a_2, 0)$ . Given  $\mu$ , and the population experimentation rate  $\varepsilon$ ,  $j$  chooses the optimal experimentation rate  $\alpha$  out of the compact set  $[0, 1]$ . Consider the joint process  $\{Y_k\}$  associated with  $\omega(m_k) = \omega_1$ . Set  $\Delta u(s_1, s_2, \omega_1) = 1$  for  $(s_1, s_2) = (a_1, 1; a_2, 0)$  and  $(s_1, s_2) = (a_1, 0; a_2, -1)$ , put  $\Delta u(s_1, s_2, \omega_2) = -1$  for  $(s_1, s_2) = (a_2, -1; a_1, 0)$  and  $(s_1, s_2) = (a_2, 0; a_1, 1)$ , and set  $\Delta u(C) = 0$ . The assignment of  $\Delta u(s_1, s_2, \omega_2)$  is analogous. It follows that:

$$\Delta U(a_2, 0 \mid \varepsilon) = E \left[ \sum_{k=0}^{\infty} \Delta u(Y_k, \omega_i) \mid (a_2, 0), \varepsilon \right].$$

Since the population law of motion  $x(t)$  is continuous in  $\varepsilon$ , and since the matching and state-change processes are independent of  $\varepsilon$ , the expression  $\Pr(Y_k, \omega_i \mid (a_2, 0), \varepsilon)$  must be continuous in  $\varepsilon$ , for any  $k, Y_k$ , and  $i$ . Thus  $\Delta U(a_2, 0 \mid \varepsilon)$  is continuous in  $\varepsilon$ . Since  $\Delta U(a_2, 0 \mid 1) < 0$  (Lemma 5.3) and  $\Delta U(a_2, 0 \mid 0) > 0$  (Lemma 5.2), by continuity and the intermediate value theorem, there must exist a  $\varepsilon \in (0, 1)$  such that  $\Delta U(a_2, 0 \mid \varepsilon) = 0$ . Since the

population strategies are symmetric, it follows that  $\Delta U(a_1, 0 | \varepsilon) = 0$ , too. This means that  $\varepsilon \in \alpha^\mu(\varepsilon) = [0, 1]$ .

Lemma 5.3 shows that for any  $\bar{\varepsilon} > 0$ , there is  $\bar{\mu}(\bar{\varepsilon})$  such that for any  $\mu \geq \bar{\mu}(\bar{\varepsilon})$ , it is the case that  $\Delta U(a_2, 0, \varepsilon) < 0$  unless  $\varepsilon < \bar{\varepsilon}$ . It follows that  $\lim_{\mu \rightarrow \infty} \sup\{\varepsilon | \varepsilon \in \alpha^\mu(\varepsilon)\} = 0$ .

*Proof of Proposition 5.4.* We proceed in two steps.

*Step 1.* Construction of the auxiliary process  $\hat{x}^\mu$ .

Let  $x_{\mu, \varepsilon}(t', x_0)$  be the solution to the Cauchy problem identified by Eq. (3) and by the initial state  $x_0$ . Let  $T_{\mu, \varepsilon}(x_0, x)$  be the time  $t'$  such that  $x_{\mu, \varepsilon}(t', x_0) = x$ ; set  $T(\varepsilon, \mu) = T_{\mu, \varepsilon}(\underline{x}_\varepsilon, 1/2)$ . For any small  $\delta > 0$ , straightforward but tedious calculations show that:

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ \mu \rightarrow \infty}} \frac{T_{\mu, \varepsilon}(\underline{x}_\varepsilon, \underline{x}_\varepsilon + \delta)}{T_{\mu, \varepsilon}(\underline{x}_\varepsilon, 1/2)} = 1 \quad \text{and} \quad \lim_{\substack{\varepsilon \rightarrow 0 \\ \mu \rightarrow \infty}} \frac{T_{\mu, \varepsilon}(\bar{x}_\varepsilon - \delta, \bar{x}_\varepsilon)}{T_{\mu, \varepsilon}(1/2, \bar{x}_\varepsilon)} = 1.$$

Consider an arbitrary  $\omega_1$ -duration  $(\tau_i, \tau_{i+1})$ . For  $\mu$  large enough,  $x(\tau_i) \approx \underline{x}_\varepsilon$  and  $x(\tau_{i+1}) \approx \bar{x}_\varepsilon$  almost surely. This means that, for  $\varepsilon$  small and  $\mu$  large, for any  $t \in (\tau_i, \tau_{i+1})$ ,

$$\Pr(x(t) - \underline{x}_\varepsilon \in (0, \delta) | x(t) \leq 1/2) \approx 1, \quad \Pr(\bar{x}_\varepsilon - x(t) \in (0, \delta) | x(t) > 1/2) \approx 1.$$

Thus for any  $t < T(\varepsilon, \mu) + \tau_i$ , we can approximate  $x(t)$  with  $\underline{x}_\varepsilon + \delta/2$ ; for any  $t > T(\varepsilon, \mu) + \tau_i$ , we can approximate  $x(t)$  with  $\bar{x}_\varepsilon - \delta/2$ . Since

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ \delta \rightarrow 0}} \frac{\underline{x}_\varepsilon + \delta/2}{\varepsilon} = 1, \quad \text{and} \quad \lim_{\substack{\varepsilon \rightarrow 0 \\ \delta \rightarrow 0}} \frac{\bar{x}_\varepsilon - \delta/2}{1 - \varepsilon} = 1,$$

we can further approximate  $x(t) \approx \varepsilon$  for any  $t < T(\varepsilon, \mu) + \tau_i$  and  $x(t) \approx 1 - \varepsilon$  for any  $t > T(\varepsilon, \mu) + \tau_i$ . On the  $\omega_2$ -duration  $(\tau_{i+1}, \tau_{i+2})$ , by symmetry, we can approximate  $x(t) \approx 1 - \varepsilon$  for any  $t < T(\varepsilon, \mu) + \tau_{i+1}$  and  $x(t) \approx \varepsilon$  for any  $t > T(\varepsilon, \mu) + \tau_{i+1}$ .

By Lemmas 5.3 and 5.2, in equilibrium it must be the case that  $\Delta U(a_2, 0 | \varepsilon) = 0$ . We rewrite this expression conditioning on the events  $\{t < T(\varepsilon, \mu)\}$  and  $\{t > T(\varepsilon, \mu)\}$  and making explicit the dependence of  $\Delta U(a_2, 0 | \varepsilon)$  on  $\mu$ :

$$T(\mu, \varepsilon) \Delta U(\mu, \varepsilon, t < T(\mu, \varepsilon)) + (1 - T(\mu, \varepsilon)) \Delta U(\mu, \varepsilon, t > T(\mu, \varepsilon)) = 0.$$

We know that  $T(\mu, \varepsilon) \rightarrow \infty$  for  $\varepsilon \rightarrow 0$  and that  $T(\mu, \varepsilon) \rightarrow 0$  almost surely for  $\mu \rightarrow \infty$ . Since we want to study the value of  $T(\mu, \varepsilon)$  for  $\varepsilon \rightarrow 0$  and  $\mu \rightarrow \infty$ , we isolate the time  $T$  and rewrite the equation above as follows:

$$0 = T \Delta U(\mu, \varepsilon, t \leq T) + (1 - T) \Delta U(\mu, \varepsilon, t > T) \quad (C.2)$$

and

$$T = T(\mu, \varepsilon). \quad (C.3)$$

*Step 2.* Approximation of  $\varepsilon(\mu)$  for  $\mu$  large and  $\varepsilon$  small.

For  $\varepsilon$  small and  $\mu$  large, one can show (calculations available upon request to the authors) that Eq. (C.2) is approximated by the following expression:

$$\left(\frac{1}{2} \mu T - \frac{1}{3} \mu^2 T^3\right) \varepsilon + \frac{1}{2} \mu T^2 - 1 \approx 0.$$

This is a cubic equation that has only one admissible solution  $T(\mu, \varepsilon) \in (0, 1)$ ; the solution is continuous in  $\varepsilon$ . Since we are interested in approximation for small  $\varepsilon$ , we first set  $\varepsilon = 0$  and solve for  $T(\mu, 0)$ . The unique admissible solution of the equation  $\frac{1}{2} \mu T(\mu, 0)^2 - 1 = 0$  is

$$T(\mu, 0) = \frac{1}{\sqrt{\mu}} \sqrt{2}.$$

By applying Dini's implicit function theorem, we can write a linear approximation for  $\varepsilon$  small:

$$T(\varepsilon, \mu) \approx \frac{\sqrt{2\mu}}{\mu} + \frac{1}{6} \varepsilon.$$

Straightforward but tedious manipulations of Eq. (C.3) yield:

$$T(\mu, \varepsilon) = \frac{\ln\left(\frac{1}{2\varepsilon(-2\varepsilon + \sqrt{(1+4\varepsilon^2)})}\right)}{\mu \sqrt{(1+4\varepsilon^2)}}.$$

Note that  $\mu T(\mu, \varepsilon) \rightarrow \infty$ , for  $\varepsilon \rightarrow 0$ . Solving the system identified by the two equations above, we obtain that for  $\varepsilon$  small and  $\mu$  large,

$$\frac{\ln\left(\frac{1}{2\varepsilon(-2\varepsilon + \sqrt{(1+4\varepsilon^2)})}\right)}{\sqrt{(1+4\varepsilon^2)}} \approx \sqrt{2\mu} + \frac{1}{6} \mu\varepsilon.$$

For  $\varepsilon \approx 0$ , the above condition is approximated by the expression:

$$\varepsilon \approx \frac{1}{2} e^{-\sqrt{2\mu}}.$$

When  $\lambda_1 \neq \lambda_2$ , one needs to expand system (C.2) so as to allow for two different transition times,  $T_1(\mu, \varepsilon)$  and  $T_2(\mu, \varepsilon)$ , associated with the states  $\omega_1$  and  $\omega_2$ . When deriving  $T_1(\mu, \varepsilon)$  and  $T_2(\mu, \varepsilon)$  from the associated law of motions, we see that  $T_1(\mu, \varepsilon)/T_2(\mu, \varepsilon) \rightarrow 1$  for  $\mu \rightarrow \infty$ . Thus the derivation of Condition (C.4) is analogous to the case where  $\lambda_1 = \lambda_2$ .

## REFERENCES

1. A. Banerjee, A simple model of herd behavior, *Quart. J. Econ.* **107** (1992), 797–817.
2. A. Banerjee and D. Fudenberg, Word of mouth learning, mimeo, MIT, 1996.
3. J. Bjornestedt, Experimentation, imitation and evolutionary dynamics, mimeo, University of Stockholm, 1993.
4. J. Bjornestedt and J. Weibull, Nash equilibrium and evolution by imitation, in “Rationality in Economics” (K. Arrow and E. Colombatto, Eds.), MacMillan, New York, 1993.
5. R. Boylan, Laws of large numbers for dynamical systems with randomly matched individuals, *J. Econ. Theory* **57** (1992), 473–504.
6. T. Borgers and R. Sarin, Learning through reinforcement and replicator dynamics, *J. Econ. Theory* **77** (1999), 1–14.
7. A. Cabrales and J. Sobel, On the limit points of discrete selection dynamics, *J. Econ. Theory* **57** (1992), 407–419.
8. M. Davis, “Markov Models and Optimization,” Chapman and Hall, London, 1993.
9. E. Dekel and S. Scotchmer, On the evolution of optimizing behavior, *J. Econ. Theory* **57** (1992), 392–406.
10. R. Durrett, “Probability Theory and Examples,” Wadsworth, Belmont, CA, 1996.
11. P. Dutta, What do discounted optima converge to?: A theory of discount rate asymptotics in economic models, *J. Econ. Theory* **55** (1991), 64–94.
12. G. Ellison and D. Fudenberg, Rules of thumb for social learning, *J. Polit. Econ.* **51** (1993), 612–643.
13. D. Friedman, Evolutionary games in economics, *Econometrica* **59** (1991), 637–666.
14. J. Hofbauer and J. Weibull, Evolutionary selection against dominated strategies, *J. Econ. Theory* **71** (1996), 558–573.
15. G. Keller and S. Rady, Optimal experimentation in a changing environment, *Rev. Econ. Stud.* **66** (1999), 475–507.
16. J. Nachbar, Evolutionary selection dynamics in games: Convergence and limit properties, *Int. J. Game Theory* **19** (1990), 59–89.
17. G. Moscarini, M. Ottaviani, and L. Smith, Social learning in a changing world, *Econ. Theory* **11** (1998), 657–665.
18. A. Rustichini and A. Wolinsky, Learning about a variable demand in the long run, *J. Econ. Dynam. Control* **19** (1995), 1283–1292.
19. L. Samuelson and J. Zhang, Evolutionary stability in asymmetric games, *J. Econ. Theory* **57** (1992), 363–391.

20. K. Schlag, Why imitate, and if so, how?, *J. Econ. Theory* **78** (1998), 130–156.
21. L. Smith and P. Sorensen, Pathological outcomes of observational learning, *Econometrica* **68** (2000), 371–398.
22. J. W. Weibull, “Evolutionary Game Theory,” MIT Press, Cambridge, MA, 1995.
23. X. Vives, Learning from the experience of others: a welfare analysis, *Games Econ. Behavior* **20** (1997), 177–200.