Imitation and Experimentation in a Changing Environment*

Francesco Squintani† Juuso Välimäki‡

October 1999

Abstract

This paper analyzes the equilibrium play in a random matching model with a changing environment. 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 mis-coordination for a positive fraction of time.

JEL classification numbers: C73, D83, D84

*We wish to thank Eddie Dekel, Jeff Ely, George Mailath, Wolfgang Pesendorfer, Karl Schlag, Lones Smith, Asher Wolinsky, and all the participants to a Math Center seminar at Northwestern University. All remaining errors are ours.

† Department of Economics, Northwestern University, Evanston, IL 60208-2009, U.S.A., squinzio@nwu.edu

‡Department of Economics, Northwestern University, Evanston, IL 60208-2009, U.S.A., valimaki@nwu.edu, and Department of Economics, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom, juuso.valimaki@soton.ac.uk

1 Introduction

When economic agents can observe sample outcomes 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 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 experimentation is sufficient to coordinate almost all the agents 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, 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 the first action, then she (and 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, independently 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 future stream of utilities. For simplicity we assume that players retain only single period histories and hence they condition only on the outcome in the previous match.[1] It is not hard to see that a symmetric adoption of the 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. 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 since the choice of a dominated action results almost certainly in the detection of the true state in the next period, and hence the play will revert to original play. A successful experiment, however, has payoff implication beyond the 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 adoption of imitating strategies cannot be an equilibrium.

[1]We will show that this bounded-rational strategy yields almost the same payoff of fully-rational strategies for our case of interest.

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 dominated 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 1993. 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 analyzes the adoption of new technologies. Section 7 concludes, and the proofs are in Appendix.

2 Related Literature

This paper is connected to three strands of literature. In the herding literature, Ellison and Fudenberg (1995) identify conditions under which players will select the "correct" action given the state of the world, when sampling from population play and adopting a "must-see-to-adopt" rule (i.e. players may change their action only if they sample some players taking a better alternative). Banerjee and Fudenberg (1996) allow players to adopt fully-

rational decision rules and show that if players sample from the population in a proportional fashion, and signals are informative enough to outweight the prior, at the only stationary outcome all agents make the correct choice. To our knowledge, this paper is the first to study experimentation and social learning in a changing environment with forward looking agents. Smith and Sorensen (1999) explicitly introduce forward-looking behavior in a fixed environment and show that the set of stationary cascades shrinks as individuals become more patient. Moscarini, Ottaviani and Smith (1998) analyze a social learning model in a changing world with myopic players, their result on the fragility of cascades depends on exogenous private signals relative to the state of the world.

The implications of sampling from population play have been studied extensively in the evolutionary games literature. A preliminary contribution is the work of Boylan (1992) that identified matching schemes that allow the approximation of the stochastic population evolution by means of a dynamic system. Nachbar (1990), Friedman (1991) and Samuelson and Zhang (1992) 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 is full support (see also Dekel and Scotchmer 1992, Cabrales and Sobel 1992, Bjornestedt 1993, and Hofbauer and Weibull 1996). Specific characterizations of payoff-monotonic dynamics have then been derived in models of 'learning by sampling the population play' by Bjornestedt (1993), Bjornestedt and Weibull (1993), Schlag (1998), and Borgers and Sarin (1999).

Models of experimentation in a changing world were treated in the single agent case by Rustichini and Wolinsky (1995) and by Keller and Rady (1999), in a setting where a monopolist needs to choose between a sure action and an uncertain alternative whose value changes randomly over time. They show that patient players will converge on the optimal action almost all the times 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 μ to play one of two symmetric 2×2 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 Δt small. The two possible payoff matrices G_1 and G_2 are given by:

G_1 :

	a_1	a_2
a_1	$(0,0)$	$(1,-1)$
a_2	$(-1,1)$	$(0,0)$

G_2 :

	a_1	a_2
a_1	$(0,0)$	$(-1,1)$
a_2	$(1,-1)$	$(0,0)$

Note that action a_i is strictly dominant in game G_i for $i = 1, 2$.[2] If the players are not sure which 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 on relative payoff comparison among individuals, and to rule any informational content of the absolute value of a player's payoff. Denote the set of player i'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 $m^j(t)$ has an interpretation as the last time before t in which j was matched. 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.

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 $\Omega = \{\omega_1, \omega_2\}$. The key ingredient in this paper is that the state changes exogenously over time. For concreteness, we assume that the state change process is another Poisson process with parameter λ.[3] Let

[2] 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.

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

$a^j(t)$ denote the action that player j would choose if matched at instant t. The evolution of play in the population is governed by the strategies of the players and the random state changes.

4 Myopic Optimization

Consistently with previous contributions in evolutionary games literature, we assume in this section that each player maximizes 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 then denoted by $\mathbf{a^j}$, and the actions taken by $j's$ opponents are denoted by $\mathbf{a^{j(t)}}$. Let $\mathbf{u_j}\left(\mathbf{a^j}, \mathbf{a^{j(t)}}\right)$ denote the vector of realized payoffs. The history observable to player j at t is $h^j(t) = \left(\mathbf{a^j}, \mathbf{a^{j(t)}}, \mathbf{t}, \mathbf{u_j}\left(\mathbf{a^j}, \mathbf{a^{j(t)}}\right), t\right)$, where the last component underlines that strategies may depend on calendar time.

A pure strategy of an arbitrary player j at instant t is then

$$s^j : h^j(t) \to \{a_1, a_2\}.$$

Denoting by $a(t)$ the random action of a player from the population at time t, we will assume that player j prefers action a_1 to action a_2 at time t if

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

The first Proposition shows that the unique equilibrium is such that players will always adopt the imitation rule.

Proposition 1 *If players are myopic optimizers, for any λ, and μ, the equilibrium strategy:*

$$\begin{aligned} s(a_1, 1) &= s(a_1, 0) = s(a_2, -1) = a_1 \\ s(a_2, 1) &= s(a_2, 0) = s(a_1, -1) = a_2. \end{aligned} \tag{1}$$

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

We denote the strategies derived in Proposition 1 as *imitation strategies.* It is important to notice that if G_i is the game played in all t, this dynamics leads to an asymptotic steady state where all players correctly assess the state of the world ω_i and play action a_i. The following analysis shows that in changing environments, imitation strategies do not allow players to correctly assess the state of the world over time.

Denote the population fraction of players using a_1 in period t by $x(t)$, i.e. 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. In ω_1, the law of motion for $x(t)$ is given (almost everywhere) by:

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

Of all the matched players (that have instantaneous flow rate of μ), 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 $x(0)$:

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

The dynamics for the population fraction playing a_2 follows immediately from the constant population assumption. A similar derivation can be done for state ω_2 to yield:

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

The main task in this section is to patch these two dynamics together to yield an overall population dynamics in the changing environment. Start the system without loss of generality at $x(0) = \frac{1}{2}$, $\omega(0) = \omega_1$. Denote the random times of state changes by $\{\tau_i\}_{i=1}^{\infty}$, where τ_i is the physical instant of i^{th} state switch. Notice that for i odd, the switches are from ω_1 to ω_2 and for i even, they are from ω_2 to ω_1. As the state changes are characterized by a Poisson arrival process, the expected waiting time between τ_i and τ_{i+1} is $\frac{1}{\lambda}$. To derive the asymptotic properties of this process, consider the population shares at the switching times, $x(\tau_i)$. We know that $x(\tau_i) < x(\tau_{j+1})$ for i even and the reverse inequality holds for i odd. Consider first the following limit:

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

where $I_A(x(t)) = 1$ if $x(t) \in A$ and $I_A(x(t)) = 0$ if $x(t) \notin A$. This limit 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 1 *If players are myopic optimizers, for any λ, and μ, then $x_A = 0$ 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 1 *State ω_i is ϵ-undetected for duration i if $x(t) \geq 1 - \epsilon$ for $\tau_i \square t \square \tau_{i+1}$, i odd and ω_1 is ϵ-undetected for duration i if $x(t) \square \epsilon$ for $\tau_i \square t \square \tau_{i+1}$, i even.*

In words, all but at most ϵ fraction of players in the population play the dominated action for the time interval in question. The previous lemma can be used to prove the following proposition.

Proposition 2 *If players are myopic optimizers, for any λ, and μ, in the long run, a half of the state durations is ϵ-undetected for all $\epsilon > 0$.*

In words, the proposition above characterizes the frequency of large shifts in the population play in relation to the frequency of the state changes. 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.

Remark 1 In case the state changes occur with different probability, or gains for taking the dominant action do not coincide, the players will not always adopt imitative strategies. Suppose in fact that the state changes from ω_j to ω_i with rate λ_i, and without loss of generality, that $\lambda_2 > \lambda_1$. Then the optimal decision rule includes histories after which a player adopts action a_2 even though at the previous match she played a_1 and tied with the opponent. The main result of the section, Proposition 2, however continues to hold in the sense that the population play concentrates on the action a_2 in the long run. Therefore the ω_1-state durations will be ϵ-undetected for all $\epsilon > 0$. The same result holds in the case that $u(a_2, a_1, \omega_2) > u(a_1, a_2, \omega_1)$. In sum, dominated actions are adopted in the long run by a majority of the population of myopic players for fractions at least $\min\left\{\frac{\lambda_1}{\lambda_1+\lambda_2}, \frac{\lambda_2}{\lambda_1+\lambda_2}\right\}$ of the total time. More details are in the Appendix.

5 Forward-Looking Optimization

In this section, we assume that each player cares about her own payoff as well as about the her current payoff as well as the future payoffs. At the same time, for analytical tractability, we assume use stationary strategies with a single period memory. We will see at the end of the section that such an assumption does not entail significant predictive loss. As in the previous section, we present only the case for $\lambda_1 = \lambda_2$.[4]

We can define the history observable to j at t as

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

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 implicitly assuming here that players do not know $m^j(t)$, i.e. the strategies do not depend on calendar time. A pure strategy of an arbitrary player j at instant t is then

$$s^j : h^j(t) \to \{a_1, a_2\}.$$

In order to simplify the calculations, we use the overtaking criterion rather than the discounted sum of payoffs for evaluating sequences of outcomes.[5] 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 $\left\{a^{j(m_k)}(m_k)\right\}_{k=0}^{\infty} \equiv \mathbf{a^{j(m)}}$. To evaluate the utilities from various action profiles, we consider the following infinite summations:

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

[4]The case for $\lambda_1 \neq \lambda_2$ yields indistinguishable results. The Appendix, when presenting the proofs of the statements presented in this section, also points out the differences for the case when $\lambda_1 \neq \lambda_2$.

[5]The limit of means criterion is does not discriminate enough bewteen sequences of outcomes for our purposes since the effect of any individual decisions is vanishing in the limit (the processes that result from the analysis are strongly mixing).

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

Consider player j at the moment of her decision between a_1 and a_2, let her choice be called a_i, and notice that a_1 and a_2 induce different distribution of continuation plays. Let the future actions conditional on an initial choice a_i by $\mathbf{a_i^j}$ so that a choice at matching instant m_k following initial choice a_i is given by $a_i^j(m_k)$. Let $\mathbf{m}$, $\mathbf{a_i^j}$ and the actions of future opponents, $\mathbf{a^{j(m)}}$, 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 $\overline{K} < \infty$ such that for all $K \geq \overline{K}$,

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

where the expectations are taken with respect to the random matching probabilities. In the last part of the section, the impact of the current choice on the future choices and payoffs is made explicit.

We solve for the symmetric stationary equilibrium strategies of the game. Formally, we are looking for a strategy $s \in S$ such that it is optimal for each player to use s if all the other players use s. Notice that here we are assuming that a player's own future choices comply with s.

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.

Proposition 3 *For any μ and λ, at equilibrium, $s(a_1, 1) = s(a_2, -1) = a_1$, and $s(a_2, 1) = s(a_1, -1) = a_2$.*

While imitation 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. In what follows, we use σ to indicate the mixed strategy of any given player.

Proposition 4 *For all λ, there is a $\mu(\lambda)$ such that, at the stationary, symmetric equilibrium,*

$$\Pr\{s(a_1,0)=a_1\}<1,\ \Pr\{s(a_2,0)=a_2\}<1$$

whenever $\mu \geq \mu(\lambda)$.

The intuition behind this result is quite simple. If all the other players are using the imitation strategies from the previous section, then it is optimal for a given individual to change her action conditional on observing a tie in her previous match. The reason for this is that the play within the population does not react to most state changes in the long run, and therefore a single trial leads to a large expected number of wins in the future if the population is currently concentrated on the inferior action. If the population is concentrated on the dominant action, a change of actions leads to a single period loss. Therefore the gains are obtained over many periods if the meeting rate is high, and losses take place in a single period, and it is optimal to change actions if μ is high enough.

Given that the state changes with the same rate from ω_1 to ω_2 and from ω_2 to ω_1, it is meaningful to restrict attention to equilibria where $\Pr\{s(a_1,0)=a_1\}=\{s(a_2,0)=a_2\}$, and we introduce $\varepsilon=1-\Pr\{s(a_i,0)=a_i\}$. In words, the relevant strategies always choose the dominant action in the previous match if it is identified by the outcome. If the state is not revealed in the previous match, then the same action, a_i, as before is chosen with probability $1-\varepsilon$.

For each fixed μ, and for any choice of the experimentation parameter ε, we can derive the law of motion for the population choices of actions. As before, let $x_\varepsilon(t)$ denote the fraction of players choosing action a_1 in period t. Note that we are parametrizing the process

of population play by the relevant experimentation probabilities. In state ω_1, we have:

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

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_ε. The relevant root of the quadratic equation is:

$$\bar{x}_{\varepsilon_1,\varepsilon_2} = \frac{1-2\varepsilon+\sqrt{1+4\varepsilon^2}}{2}.$$

The same reasoning leads to the following law of motion and the corresponding long run steady state in state ω_2

$$\begin{aligned}\dot{x}_\varepsilon(t) &= \mu\left[\varepsilon - x_\varepsilon(t)(1+2\varepsilon) + x_{\varepsilon_1,\varepsilon_2}(t)^2\right], \text{ and} \\ \underline{x}_{\varepsilon_1,\varepsilon_2} &= \frac{1+2\varepsilon-\sqrt{1-4\varepsilon^2}}{2}.\end{aligned}$$

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 $(\underline{x}_\varepsilon, \bar{x}_\varepsilon)$.[6] 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 ω_1, the rate of increase of $x(t)$ approaches 0 as $x(t)$ converges to $\bar{x}_\varepsilon$. On the other hand, the rate of decrease (i.e. also the rate of increase of action a_2) at $\bar{x}_\varepsilon$ is bounded away from zero for all $\varepsilon > 0$.[7]

[6] 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 (1996).

[7] 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.

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 ω_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 $(\underline{x}_\varepsilon, \bar{x}_\varepsilon) \times \Omega$, 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 during 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

To address the optimal choice of an action by any given player, we use a statistical technique called *coupling*. The idea is to generate two independent copies of a random process on the same probability space and deduce payoff consequences from the joint evolution. The key observation for this analysis 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 independent of the past actions of the player. As in all equilibrium analysis, each player takes the actions of all other players (including her own future actions) as given and chooses a best response. The current choice of a player has implications for her future payoffs only through the observations that are generated by that choice. The following proposition however states formally that the difference in the distribution of the continuation play induced by a different initial action choice vanishes in finite time.

Lemma 2 *For almost all* $\mathbf{m}$, $\mathbf{a}_i^{\mathbf{j}}$ *and* $\mathbf{a}^{\mathbf{j}(\mathbf{m})}$, *there exists a* $K < \infty$ *such that* $a_1^j(m_k) = a_2^j(m_k)$, *for all* $k > K$. *Furthermore,* $EK < \infty$, *where the expectation is taken with respect to the distribution of* $\mathbf{m}, \mathbf{a}_i^{\mathbf{j}}$ *and* $\mathbf{a}^{\mathbf{j}(\mathbf{m})}$.

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 under the assumption $\varepsilon_1 = \varepsilon_2 = \varepsilon$, any proposed symmetric stationary equilibrium profile is characterized by

a single parameter, ε. The exogenous environment is parametrized by (λ, μ). We hold λ fixed throughout the discussion and let μ vary. This is without loss of generality since the model with parameters $(p\lambda, p\mu)$ is equivalent to (λ, μ) apart from a linear scaling in the units of measurement for time. Fix a particular player, j, and denote her set of optimal experimentation probabilities when all others experiment at rate ε, and the rate of matches is μ by $\alpha^{\mu}(\varepsilon)$.

Lemma 3 *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 4 *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.

Since the payoff function of player j is continuous in the population experimentation rate ε since it is a time integral of a payoff that is continuous in ε against a Poisson arrival process, Lemma 3, Lemma 4 and a simple application of the intermediate value theorem allow us to conclude the main existence result of this section.

Proposition 5 *For all μ, there is an $\varepsilon > 0$ such that $\varepsilon \in \alpha^{\mu}(\varepsilon)$. Furthermore, $\lim_{\mu \to \infty} \underline{\varepsilon}(\mu) = 0$, 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 μ, the equilibrium experimentation probabilities are small. The remainder of this section investigates the asymptotic rate at which ε converges to zero as μ

increases. This is exercise is essential if we want to get a good idea of 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 μ. 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 δ to $1-\delta$ is extremely short for μ large. As a result, for μ large, $x(t)$ 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 $\mu \to \infty$, 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 $\frac{1}{2}$. 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 $\mu \to \infty$.

Proposition 6 *For any $\varepsilon(\mu)$ such that $\varepsilon(\mu) \in \alpha^{\mu}(\varepsilon(\mu))$,*

$$\lim_{\mu \to \infty} \mu T(\mu, \varepsilon(\mu)) = O(\sqrt{\mu})$$
$$\lim_{\mu \to \infty} \varepsilon(\mu) \approx \frac{1}{2} e^{\sqrt{2\mu}}$$

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 μ, and since the number of matches before the state change is $\frac{1}{2}-$detected in the terminology of the previous section (and also $1-\gamma$ detected for any $\gamma > 0$) grows linearly in $\sqrt{\mu}$, almost all the players are choosing the dominant action almost all of the time when

$\mu \to \infty$. 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 out an argument to show that it is not in any player's interest to buy costly information about past matches, and to keep track of calendar time. 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_1 if and only if $\omega(t) = 1$. Consider the expected loss of the bounded memory strategy along a renewal cycle (τ_k, τ_{k+1}), where $\omega(t) = 1$ for $t \in (\tau_k, \tau_{k+1})$. Consider T such that. $x(T + \tau_k) = 1 - \bar{x}_\varepsilon - \delta$. For each fixed δ, we know that $T \to 0$ for $\mu \to \infty$. Again set $\lambda = 1$. Against the 1-period-memory equilibrium population, the optimal strategy average payoff per renewal cycle is bounded above by $[T + (\bar{x}_\varepsilon + \delta)(1 - T)]$.

We know that $\bar{x}_\varepsilon \approx \varepsilon \to 0$ so the limiting payoff from using the optimal strategy is bounded from above by δ. 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 $\delta - C$.

Consider the average payoff per state duration in ω_1 when using the equilibrium strategy. By revealed preference we know that such a payoff is not larger than the average payoff obtained by a player using the pure imitation strategies of the second section. 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 $\delta > 0$ small enough and μ large enough to have $-\frac{1}{\mu} > \delta - C$.

If information about past moves and calendar time came for free, the optimal strategy of player j would be summarized by a sequence of switching times $\{T_k\}_{k=1}^{\infty}$ we call *alarm-clocks*. Letting T_0 denote the time of the last matching revealing the state of the world, player j will play the pure-imitation action at times $t \in (T_k, T_{k+1})$ k even, and the opposite action at times $t \in (T_k, T_{k+1})$. The optimal sequence $\{T_k\}_{k=1}^{\infty}$ represents the instants in which the expected differential payoff for reverting action crosses a threshold derived from the indifference principle. Since non-revealing matches such as $(a_1, 0)$ are informative of the

population play the optimal sequence does not only depend on the time of the last state-revealing match, but also on the instants of subsequent non-revealing matches. Because of that, the exact determination of optimal alarm-clocks becomes rather messy.

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 own past choices regardless of the actions that other agents chose. The techniques developed here would be useful for those models as well as long as social learning is not swamped by learning from 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.

7 Appendix

7.1 Proofs Omitted from Section 4

Proof of Proposition 1. Since players are myopic, the benefit for successful experimentation is $+1$, and the loss for unsuccessful experimentation is -1. Without loss of generality, set equal to 0 the time of the last match that revealed the state of the world. Thus a player's payoff depends only on whether an odd or even number k of renewals has occurred since time $t = 0$. For any $x \geq t$,the renewal equation is simply,

$$\begin{aligned} \Pr(0;(x,t),1) &= e^{-\lambda(t-x)} \\ \Pr(k;(x,t),1) &= \int_x^t \lambda e^{-\lambda s} \Pr(k-1;(s,t),2)ds \end{aligned}$$

iteratively for any $k > 0$. Recursive calculations show that for any $t > 0$, and any $l \geq 0$, $\Pr(2l; (0,t), 1) > \Pr(2l+1; (0,t), 1)$. Therefore the optimal strategy is always imitation. ■

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

Define the sequence of random variables $\{\sigma_k\}$, $\{y_{1k}\}$, and $\{\beta_k\}$ as follows. Let $\Delta_{1s} = \tau_{2s+1} - \tau_{2s}$ and $\Delta_{2s} = \tau_{2s} - \tau_{2s-1}$. Then for $i = 1, 2$, $\Delta_{is} \sim \exp\left(\frac{1}{\lambda_i}\right)$, i.i.d.

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

$\{\sigma_k\}$ is a random walk with a strictly positive, but bounded variance. Then $\beta_k > \sigma_k$ and $\sigma_{k+1} < \beta_k$. Notice that β_k is also a martingale for $k \geq 1$. It is easy to check that

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

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

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

whenever $\sigma_k > K(\epsilon)$. Note that this also yields $x(\tau_{2k+1}) < \epsilon$ whenever $\beta_k < -K(\epsilon)$.

Since both σ_k and β_k are driftless random walks with strictly positive (and constant) variance for all k, the expected time to re-entry to $[-\infty, K(\epsilon)]$ by σ_k is infinite (and similarly for β_k) by the contrapositive of the Wald equation (see Durrett 1996). Starting inside $[-K(\epsilon), K(\epsilon)]$, the expected hitting time for σ_k (and for β_k) to $[K(\epsilon), \infty] \cup [-\infty, -K(\epsilon)]$ is bounded above by $K(\epsilon)^2 / E[y_1^2] < \infty$. In fact $\sigma_k^2 - kE[y_1^2]$ is a martingale, and letting $T = \inf\{k : \sigma_k \notin [-K(\epsilon), K(\epsilon)]\}$, and starting with $\sigma_k^2 = 0$ it follows that $K(\epsilon)^2 = E[\sigma_T^2] = E[T]E[y_1^2]$.

Thus we know that

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

■

Proof of Proposition 2. We need to show that for all $\epsilon > 0$,

$$\lim_{i\to\infty} \frac{\#\{i \text{ odd} : x(t) \geq 1-\epsilon \text{ for } \tau_i \square\, t \square\, \tau_{i+1}\} + \#\{i \text{ even}: x(t) \square\, \epsilon \text{ for } \tau_i \square\, t \square\, \tau_{i+1}\}}{i} = \frac{1}{2}.$$

By the previous result, choose $K(\epsilon)$ to be such that $\beta_k > K(\epsilon) \Rightarrow x(\beta_k) > 1-\epsilon.$, 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 $x_{[\epsilon, 1-\epsilon]} = 0$. ■

7.2 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 correspondent on the state with highest flow rate, so that all durations in the smallest flow state will be undetected.

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

First note that if λ_1 is very dissimilar to λ_2, the optimal strategy cannot be imitation. It is in fact straightforward to notice that $\lambda_1/\lambda_2 \to \infty$, and μ 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 ω_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 with history $(a_1, 0)$, and let τ denote the last time j was matched. For τ large enough, since the process is strongly mixing, the relative probability of the state being ω_1 will approximate $\lambda_1/[\lambda_1 + \lambda_2]$, and thus player j will play action a_2.

On the other hand, if $\lambda_2 > \lambda_1$, a straightforward extension of the proof of Proposition 1 yields that at equilibrium $s(a_1, -1) = s(a_2, 0) = s(a_2, 1) = a_2$: players never abandon action a_2 unless it was defeated at the previous match.

In the next Proposition we will show that if $\lambda_2 > \lambda_1$, the ω_1-state durations will not be detected by the population dynamics induced by pure imitation. The result holds a fortiori for the dynamics induced by equilibrium strategies, because, under the latter, a_1 is played after a non-larger set of histories than implied pure imitation.

Proposition 7 *If $\lambda_1 < \lambda_2$, then $x(t)$ converges to 1 (almost surely) as t goes to infinity. If $\lambda_1 > \lambda_2$ then x_t converges to 0 (almost surely) as t goes to infinity.*

Proof. The sequence $\{\sigma_k\}$ is a random walk with a strictly positive, but bounded variance. The recurrence properties of this walk depend on whether y_s has a zero mean. If the mean is positive, i.e., $\lambda_1 < \lambda_2$, then by strong law of large numbers, for all K, $\Pr\{\sigma_k < K \text{ for infinitely many } k\} = 0$. It is easy to check that

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

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

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

Since $\sigma_k > K(\varepsilon)$ for all but finitely many k, the almost sure convergence of the process to 1 follows. A similar construction applies to the case where $\lambda_1 > \lambda_2$. ■

An easy way of understanding the intuition behind this result is to notice the symmetry of equations 2 and 3. Hence we could interpret the two imitation processes as motions along the same curve in different directions. To find out what happens to the position of a particle, we need to look at the combined impact of the two processes on the total displacement. But this is obtained simply by multiplying the speed by expected time to next state change, and the result follows.

7.3 Proof Omitted from Section 5

Proof of Proposition 3. Since the player knows the state of the world at the last meeting, her optimal choice will depend only on the probability of the next meeting m to occur exactly after k and before $k+1$ state-switches. Let that event be: $\{\tau_k < m < \tau_{k+1}\}$. The renewal system is as follows:

$$\begin{cases} \Pr\{\tau_k < m < \tau_{k+1}\} = \int_0^\infty \Pr\{\tau_{k-1} < m - t < \tau_k\} \lambda e^{-\lambda t} dt & \text{if k>0} \\ \Pr\{\tau_0 < m < \tau_1\} = \int_0^\infty \left[\int_s^\infty \lambda_1 e^{-\lambda t} dt\right] \mu e^{-\mu s} ds = \frac{\mu}{\mu+\lambda} \end{cases}$$

By induction, we obtain:

$$\Pr\{\tau_k < m < \tau_{k+1}\} = \frac{\mu \lambda^k}{\prod_{t=1}^{k+1} [\mu + t\lambda]}.$$

Since $\Pr\{\tau_k < m < \tau_{k+1}\} = \Pr\{\tau_{k-1} < m < \tau_k\} \frac{\lambda}{\mu + (k+1)\lambda}$, it follows that $\Pr\{\tau_k < m < \tau_{k+1}\}$ is strictly decreasing in k.

Then, for any k odd, $\Pr\{\tau_k < m < \tau_{k+1}\} < \Pr\{\tau_{k-1} < m < \tau_k\}$. So that

$$\Pr\{\omega_m = 2|\omega_1\} = \sum_{k=0}^{\infty} \Pr\{\tau_{2k+1} < m < \tau_{2k+2}\} < \sum_{k=0}^{\infty} \Pr\{\tau_{2k} < m < \tau_{2k+1}\} = \Pr\{\omega_m = 1|\omega_1\}.$$

Since the payoffs are symmetric, the conclusion is that the optimal strategy is imitation, regardless of μ.

In case with $\lambda_1 \neq \lambda_2$, since the player knows the state of the world at the last meeting, her optimal choice will depend only on the probability of the next meeting m to occur

exactly after k and before $k+1$ state-switches, when the state at the current meeting is ω_1 (the analysis for ω_2 is symmetric). Let that event be: $\{\tau_k < m < \tau_{k+1}|\omega_1\}$. The renewal system is as follows:

$$\begin{cases} \Pr\{\tau_k < m < \tau_{k+1}|\omega_1\} = \int_0^\infty \Pr\{\tau_{k-1} < m-t < \tau_k|\omega_2\}\lambda_1 e^{-\lambda_1 t}dt & \text{if k>0} \\ \Pr\{\tau_0 < m < \tau_1|\omega_1\} = \int_0^\infty \left[\int_s^\infty \lambda_1 e^{-\lambda_1 t}dt\right]\mu e^{-\mu s}ds = \frac{\mu}{\mu+\lambda_1} \end{cases}$$

By induction, we obtain that

$$\Pr\{\tau_k < m < \tau_{k+1}|\omega_1\} = \begin{cases} \frac{\mu\lambda_1^{k-[k/2]}\lambda_2^{[k/2]}}{\prod_{t=1}^k[\mu+[t/2]\lambda_1+(t-[t/2])\lambda_2]} & \text{if k is odd} \\ \frac{\mu\lambda_1^{[k/2]}\lambda_2^{k-[k/2]}}{\prod_{t=1}^k[\mu+(t-[t/2])\lambda_1+[t/2]\lambda_2]} & \text{if k is even} \end{cases}$$

As $\Pr\{\tau_k < m < \tau_{k+1}|\omega_1\}$ is continuous in $\mu, \lambda_1, \lambda_2$, it follows that there is a $M(\mu)$ s.t. if $|\lambda_1 - \lambda_2| < M(\mu)$, the optimal strategy is imitation, and $M(\mu) > 0, \forall\mu$.

The payoff for playing the imitation strategy is strictly increasing in

$$\sum_{s=0}^{\infty}[\Pr\{\tau_{2s} < m < \tau_{2s+1}|\omega_1\} - \Pr\{\tau_{2s+1} < m < \tau_{2s+2}|\omega_1\}]$$

which is increasing in μ. So it follows that $M(\mu)$ is increasing in μ.

Finally: since $\Pr\{\tau_0 < m < \tau_1|\omega_1\} \to 1$ for $\mu \to \infty$ and any fixed λ_1 and λ_2, and the optimal strategy is imitation when $\Pr\{\tau_0 < m < \tau_1|\omega_1\} > 1/2$, it easily follows that $M(\mu) \to \infty$ for $\mu \to \infty$.

For $\lambda_1/\lambda_2 \to \infty$, and μ fixed, however, $\Pr\{\tau_1 < m < \tau_2|\omega_1\} \to 1$, so that the optimal strategy is a_2: the player will not choose imitation. ■

Proof of Proposition 4. Consider first the case for $\lambda_1 \neq \lambda_2$, and say that $\lambda_1 > \lambda_2$. Proceed by contradiction, and suppose that (pure-strategies) imitation is an equilibrium. In the previous section we proved that, for any $\epsilon > 0$, under imitation strategies, eventually, $x(t) \in (0, \epsilon)$.

Consider a triple of consecutive state renewals: $\{\tau_{k-1}, \tau_k, \tau_{k+1}\}$. So that $\omega = 1$, on (τ_{k-1}, τ_k) and the population play is $x(t) \approx 0, \forall t \in [\tau_k, \tau_{k+1}]$.

Once experimented at the encounter $m \in (\tau_{k-1}, \tau_k)$, the player will know that the true state is ω_1 and play a_1 until her first encounter $m_1 \in (\tau_k, \tau_{k+1})$. As $x(t) \in (0, \epsilon)$, at m_1, she will detect the state change with probability close to 1. Thus her expected gain from experimenting is approximately $\mu[\tau_k - m]$. If she experiments at $m \in (\tau_k, \tau_{k+1})$, instead, she will lose only a payoff of 1. Solving a simple compound renewal (Poisson) process, we obtain that $\Pr(m \in (\tau_k, \tau_{k+1})) = \lambda_1/[\lambda_1 + \lambda_2]$ and $\Pr(m \in (\tau_{k-1}, \tau_k)) = \lambda_2/[\lambda_1 + \lambda_2]$ independently of μ. Since matchings and state renewals are independent processes, her

average net gain on (τ_{k-1}, τ_{k+1}) from experimenting with probability α is approximately:

$$\alpha \left[\frac{\lambda_2}{\lambda_1 + \lambda_2} \int_{(\tau_{k-1}, \tau_k)} \frac{\mu[\tau_k - m]}{\tau_k - \tau_{k-1}} dm - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right]$$

which is strictly larger than zero for μ large enough.

For $\lambda_1 = \lambda_2$, it follows that $\lambda_1/[\lambda_1 + \lambda_2] = 1/2 = \lambda_2/[\lambda_1 + \lambda_2]$, that $x(t) \in \{(0, \varepsilon), (1 - \varepsilon, 1)\}$ almost always, and that $1/2$ of the state durations are ϵ-undetected, for any $\epsilon > 0$. The above argument holds unchanged for ϵ-undetected durations. When a renewal is detected by the population play, the player expected gain from experimenting with probability α is bounded below by -1. Compounding ϵ-detected with ϵ-undetected durations, it follows that the future average net gain from experimenting with probability α is larger than $\mu^2/16 - 3/4$.

Note that the above derivation implies that if the population is playing the pure imitation strategy, each single player at each match will prefer to deviate and experiment with probability 1. ■

Proof of Lemma 2. Each player j will take an experimentation decision at matching time t either with history $(a_1, 0)$ or $(a_2, 0)$.

Step 1: *Construction of the joint process.*

Consider first the case in which the true state of the world is ω_1, and $h^j(t) = (a_2, 0)$. Consider the sequence of matching times $\mathbf{m} = \{m_k\}_{k=1}^{\infty}$: $m_k > t$, and of the state-switches $\tau = \{\tau_k\}_{k=1}^{\infty} : \tau_k > t$, for any matching time m_k, denote by x_k the fraction of the population playing a_1 at time m_k.

The realizations of $\mathbf{m}$ and τ are independent of whether j takes action a_1 or a_2. For any fixed population strategy identified by ε experimentation level, the actions of the $j's$ opponents at future matches are independent of a_i. Conditional on a_i, the transition of j opponents are identically distributed.

As long as a state-switch has not occurred, the players strategies can be described as a Markov chain with states $S = \{(a_1, 0), (a_2, 0), (a_1, 1), (a_2, -1)\}$. We proceed by pairing different states representing different continuations that depend on whether player j experimented or not. Specifically, we construct a non-autonomous Markov process as follows. Consider state space be S^2 where the first component of the couple refers to states that belong to continuation generated by j taking $a_i = a_1$, and the second component refers to states relative to $a_i = a_2$. Define by *coupling* event the union of the states such that their equilibrium strategy is the same. Formally, let $C = \{(s_1, s_2) | \Pr(a_1|s_1) = \Pr(a_1|s_2)\}$. Conditional on the event C, the distribution on the continuations given $a_i = a_1$ is identical to the distribution on the continuations given $a_i = a_2$. The coupled process is thus summarized by a 5-state non-autonomous Markov transition 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,

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

The initial distribution of the coupling process for $h^j(t) = (a_2, 0)$ is $(a_1, 1; a_2, 0)$ with probability $1 - x(t)$, and $(a_1, 0; a_2, -1)$ with probability $x(t)$.

The same transition matrix above describes the coupled process when $h^j(t) = (a_1, 0)$, as long as the first component is now meant to represent j taking $a_i = a_2$, and the second, j taking $a_i = a_1$. The two cases are subsumed by saying that the first component of the duplicated process refers to instances following (pure-strategies) experimentation and the second component, to instances following imitation. The initial distribution of the coupling process for $h^j(t) = (a_1, 0)$ is thus $(a_2, 0; a_1, 1)$ with probability $1 - x(t)$, and $(a_2, -1; a_1, 0)$ with probability $x(t)$.

In a completely analogous fashion, the transition is constructed for the case that $\omega = 2$. The probability of the next matching to occur after a state ω renewal has been calculated in a previous Lemma, it is shown to be positive. By expanding the state space to account for the transition under both ω_1 and ω_2, and compounding the relative transition probabilities we obtain a non-autonomous Markov Process we denote by $\{X_k\}_{k=1}^{\infty}$.

Step 2: *The expected time of coupling is finite.*

Looking at the above matrix, one can appreciate the key property of $\{X_k\}_{k=1}^{\infty}$: for any state $s \neq C$, the transition probability to C, $p(s, C)$ is either ε or $1 - \varepsilon$ (i.e. the process is not decomposable), and $p(C, C) = 1$ (i.e. C is absorbing). When $\varepsilon \in (0, 1)$, as that value is independent of time, we can treat the issue of recurrence of C as if we were dealing with an autonomous Markov process, to immediately conclude that, for any $s \neq C$, $P_s(T_C < \infty) = 1$, where $T_C = \inf\{k | X_k = C\}$ and P_s is the probability induced on the process by $X_m = s$.

The case for $\varepsilon = 0$ has already been ruled out from equilibrium analysis. For the case for $\varepsilon = 1$, it follows that $p((a_1, 1; a_2, 0), C) = p((a_2, 0; a_1, 1), C) = 1$. Since $\forall k$, $x_k < \bar{x}_1 < 1$, $p((a_1, 0; a_2, -1), (a_2, 0; a_1, 1) | m_k) = p((a_2, -1; a_1, 0), (a_1, 1; a_2, 0) | m_k) = (1 - x_k) > 0$. Thus the same result derived for $\varepsilon \in (0, 1)$ obtains. ■

Proof of Lemma 3. Already shown in the Proof of Proposition 4. ■

Proof of Lemma 4. Fix any $\varepsilon \geq \bar{\varepsilon}$. Consider any meeting where an individual j holds history $h = (a_2, 0)$. Let $\mathbf{x} = \{x_0, x_1, \cdots, \}$ be an arbitrary *increasing* sequence of

population plays at $j's$ future meetings, until the next ω renewal. Clearly x_0 denotes the population play at the present meeting, x_1 the population play at the next meeting and so on. Let M be the cardinality of $\mathbf{x}$.

Step 1: *The net gain for experimentation $\Delta U(a_2, 0|\omega_1, x, \varepsilon)$, when $h = (a_2, 0)$, $\omega = 1$, and the population experimentation is ε, admits a finite upper bound, independently of x.*

Consider the Markov transition derived in the proof of Proposition 2.

With probability x_0, j's opponent plays a_1. In such case, the process starts at $(a_1, 0; a_2, -1)$, player j gains a payoff of 1 from experimentation, and the path exits with probability $1-\varepsilon$. The net gain if it will never enter in $(a_1, 1; a_2, 0)$ is at most 0, and with probability smaller than $1-(1-\varepsilon)-\varepsilon(1-\varepsilon)$ it will eventually reach $(a_1, 1; a_2, 0)$. Thus

$$\begin{aligned}\Delta U(a_1, 0; a_2, -1|\mathbf{x}, \varepsilon) &< 1 + [1-(1-\varepsilon)-\varepsilon(1-\varepsilon)]\Delta U(a_1, 1; a_2, 0|\mathbf{x} - \{x_0\}, \varepsilon) \\ &< 1 + \varepsilon^2 \sum_{k=0}^{M-1} (1-\varepsilon)^k < 1 + \varepsilon\end{aligned}$$

Putting both cases together we obtain

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

the key observation is that the bound we have obtained is uniform in $\mathbf{x}$.

Step 2: *The net loss $\Delta U(a_1, 0|\omega_1, x_0, x, \varepsilon)$ admits a strictly positive bound.*

With probability x_0, j's opponent plays a_1. In such case, the process starts at $(a_2, -1; a_1, 0)$, player j's net gain is -1, the process exits with probability $1-\varepsilon$. With probability ε, the process enters the state $(a_1, 1; a_2, 0)$, with probability $1-x_1$ and the state the state $(a_1, 0; a_2, -1)$, with probability x_1. Since this is exactly the event $(a_2, 0|\omega_1, \mathbf{x} \setminus \{x_0\}, \varepsilon)$, her net gain is bounded as:

$$\begin{aligned}&\Delta U(a_2, -1; a_1, 0|\mathbf{x}, \varepsilon) \\ < \; &-1 + \varepsilon \Delta U(a_2, 0|\omega_2, \mathbf{x} \setminus \{x_0\}, \varepsilon) = -1 + \varepsilon[x_1 \frac{1}{\varepsilon} + (1-x_1)(1+\varepsilon)] = (1-x_1)(\varepsilon^2 + \varepsilon - 1)\end{aligned}$$

With probability $(1-x_0)$, j's opponent plays a_2, and the process starts at $(a_2, 0; a_1, 1)$. Player j's net gain is -1, the process exits with probability ε, and with probability less than $1-\varepsilon$, the process enters the state $(a_2, -1; a_1, 0)$. The player cannot make any positive gains unless the latter event occurs. The probability that the process loops inside $(a_2, 0; a_1, 1)$ is $(1-\varepsilon)(1-x_1)$. So

$$\begin{aligned}&\Delta U(a_2, 0; a_1, 1|\mathbf{x}, \varepsilon) < -1 + (1-\varepsilon)\Delta U(a_2, -1; a_1, 0|\mathbf{x}, \varepsilon) \\ = \; &-1 + (1-\varepsilon)\{-1 + \varepsilon[x_1 \frac{1}{\varepsilon} + (1-x_1)(1+\varepsilon)]\} = (x_1 - 1)(\varepsilon^3 + -2\varepsilon - 1) - 1.\end{aligned}$$

Wrapping up the two subcases:

$$\begin{aligned}
& \Delta U(a_1,0|\omega_2,\mathbf{x},\varepsilon) \\
< \;& x_0[-1+\varepsilon[x_1\frac{1}{\varepsilon}+(1-x_1)(1+\varepsilon)]]+(1-x_0)[-1+\varepsilon[-1+\varepsilon[x_1\frac{1}{\varepsilon}+(1-x_1)(1+\varepsilon)]]] \\
= \;& (-1+x_1+x_0-x_1x_0\,(m))\,\varepsilon^3+(-x_0x_1+x_0)\,\varepsilon^2+(x_0x_1-x_0+2-2x_1)\,\varepsilon+x_0-2+x_1 \\
= \;& p(x_1,x_0,\varepsilon)
\end{aligned}$$

Direct calculations show that $p(\bar{x}_\varepsilon,\bar{x}_\varepsilon,\varepsilon)<0$ uniformly in ε, so there must exist $\delta>0$ such that $p(x_1,x_0,\varepsilon)<0$ for $x_0>\bar{x}_\varepsilon-\delta$ and $x_1>\bar{x}_\varepsilon-\delta$. Note that since $\mathbf{x}$ is increasing, we reduce the condition to $x_0>\bar{x}_\varepsilon-\delta$.

So we conclude that $\Delta U(a_1,0|\omega_1,\mathbf{x},\varepsilon)>\hat{\Delta}>0$, which holds for any $\mathbf{x}$ s.t. $x_0>\bar{x}_\varepsilon-\delta_0$.

Step 3: *When the difference in utility $\Delta U(a_2,0)$ is expanded conditioning on the events $\{x_0<\bar{x}_\varepsilon-\delta_0\}$, and $\{x_0>\bar{x}_\varepsilon-\delta_0\}$, the first term vanishes.*

For any $\mathbf{x}$, denote by $1-\mathbf{x}$ the sequence $\{1-x_k\}$ where $x_k\in\mathbf{x}$. Using the transition matrix, and keeping in mind that payoffs are symmetric, we note that, conditional on the sequence $\mathbf{x}$ the following relationships hold

$$\begin{cases}
\Delta U(a_2,0|\omega_2,1-\mathbf{x},)=\Delta U(a_1,0|\omega_1,\mathbf{x}) \\
\Delta U(a_2,0|\omega_2,1-\mathbf{x})=-\Delta U(a_2,0|\omega_1,\mathbf{x}) \\
\Delta U(a_1,0|\omega_2,1-\mathbf{x})=-\Delta U(a_1,0|\omega_1,\mathbf{x})
\end{cases}$$

Now we can exploit the upper and lower bound, and the symmetry expanding $\Delta U(a_2,0,\varepsilon)$ as follows (for notational ease, we shall drop ε from the formula).

$$\begin{aligned}
& \Delta U(a_2,0)=\Delta U(a_2,0|\omega_1)\Pr(\omega_1|a_2,0)+\Delta U(a_2,0|\omega_2)\Pr(\omega_2|a_2,0) \qquad (4) \\
= \;& \int_{\{\mathbf{x}|x_0<\bar{x}_\varepsilon-\delta_0\}}[\Delta U(a_2,0|\omega_1,\mathbf{x})\Pr(\omega_1|(a_2,0),\mathbf{x})+\Delta U(a_2,0|\omega_2,\mathbf{x})\Pr(\omega_2|(a_2,0),\mathbf{x})] \\
& \cdot d\Pr(\mathbf{x}|x_0<\bar{x}_\varepsilon-\delta_0)\Pr(x_0<\bar{x}_\varepsilon-\delta_0) \\
+ \;& \int_{\{\mathbf{x}|x_0>\bar{x}_\varepsilon-\delta_0\}}[\Delta U(a_2,0|\omega_1,\mathbf{x})\Pr(\omega_1|(a_2,0),\mathbf{x})+\Delta U(a_2,0|\omega_2,\mathbf{x})\Pr(\omega_2|(a_2,0),\mathbf{x})] \\
& \cdot d\Pr(\mathbf{x}|x_0>\bar{x}_\varepsilon-\delta_0)\Pr(x_0>\bar{x}_\varepsilon-\delta_0) \\
= \;& \int_{\{\mathbf{x}|x_0<\bar{x}_\varepsilon-\delta_0\}}\{\Delta U(a_2,0|\omega_1,\mathbf{x})[\Pr(\omega_1|(a_2,0),\mathbf{x})-\Pr(\omega_2|(a_2,0),1-\mathbf{x})]\} \\
& d\Pr(\mathbf{x}|x_0<\bar{x}_\varepsilon-\delta_0)\cdot\Pr(x_0<\bar{x}_\varepsilon-\delta_0) \\
+ \;& \int_{\{\mathbf{x}|x_0>\bar{x}_\varepsilon-\delta_0\}}\{\Delta U(a_2,0|\omega_1,\mathbf{x})[\Pr(\omega_1|(a_2,0),\mathbf{x})-\Pr(\omega_2|(a_2,0),1-\mathbf{x})]\} \\
& d\Pr(\mathbf{x}|x_0>\bar{x}_\varepsilon-\delta_0)\cdot\Pr(x_0>\bar{x}_\varepsilon-\delta_0)
\end{aligned}$$

Consider the first term in the above equation. Let $T(\delta,\varepsilon,\mu)$ be the time taken by the population play to reach $\bar{x}_\varepsilon-\delta$, starting from $x(0)=\underline{x}_\varepsilon$, when ω_1. Solving from the

law of motion, we know that $T(\delta, \varepsilon, \mu) \to 0$ for $\mu \to \infty$. Also we know that the law of motion is increasing, which motivates our restriction to increasing $\mathbf{x}$. We know that for any time τ when the state switch from ω_2 to ω_2, the population play $x(\tau) > \underline{x}_\varepsilon$ therefore the time $\hat{T}(\delta, \varepsilon, \mu, \tau)$, taken by the population play to reach $\bar{x}_\varepsilon - \delta$, starting from $x(\tau)$ when ω_1, satisfies $\hat{T}(\delta, \varepsilon, \mu, \tau) < T(\delta, \varepsilon, \mu)$. Since the expected time until the next renewal is $1/\lambda_1 > 0$ one concludes that $\Pr(x_0 < \bar{x}_\varepsilon - \delta_0) \to 0$ for $\mu \to \infty$.

Consider now the second term of Equation (4). Introduce the likelihood ratio:

$$\Lambda = \frac{\Pr(\omega_1|(a_2, 0), \mathbf{x})}{\Pr(\omega_2|(a_2, 0), 1 - \mathbf{x})} = \frac{\Pr(\omega_1, (a_2, 0), \mathbf{x})}{\Pr(\omega = 2, (a_2, 0), 1 - \mathbf{x})}$$

If we can show that $\Pr(\omega_2, (a_2, 0), 1 - \mathbf{x}) \approx \Pr(\omega_1, (a_1, 0), \mathbf{x})$, since both are bounded away from zero, we obtain that

$$\Lambda \approx \frac{\Pr(\omega_1, (a_2, 0), \mathbf{x})}{\Pr(\omega = 1, (a_1, 0), 1 - \mathbf{x})} = \frac{\Pr(a_2, 0|\omega_1, \mathbf{x})}{\Pr(a_1, 0|\omega_1, 1 - \mathbf{x})} = \frac{1 - x_0}{x_0} < 1 - b$$

for some positive bound b. In which case, since $-\infty < \hat{\Delta} < \Delta U(a_2, 0|\omega_1, \mathbf{x}) < \bar{\Delta} < \infty$, we conclude that the second term is negative of the equation and dominates the first term, for μ large enough. That concludes that $\Delta U(a_2, 0, \varepsilon) < 0$. Invoking the symmetry already pointed out, we also note that $\Delta U(a_2, 0, \varepsilon) = \Delta U(a_1, 0, \varepsilon)$.

In Proposition 4 we showed that $P(T_C < \infty) = 1$, $p(C, C) =$ and $\Delta U(a_2, 0, \varepsilon) = 0$ conditional on C. Therefore $\Delta U(a_2, 0, \varepsilon) < 0$ implies that

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

where K' is the random match such that $m_{K'} = T_{C.}$ The overtaking criterion then implies $\alpha(\varepsilon) = 0$.

Step 4: $\Pr(\omega_2, (a_2, 0), 1 - \mathbf{x}) \approx \Pr(\omega_1, (a_1, 0), \mathbf{x})$.

Since $\Pr(\omega_2) = \Pr(\omega_1)$, we only need to show that $\Pr(a_2, 0, 1 - \mathbf{x}|\omega_2) = \Pr(1 - \mathbf{x}|\omega_2) = \Pr(\mathbf{x}|\omega_1) = \Pr(a_1, 0, \mathbf{x}|\omega_1)$.

Consider a triple of state-switch times $(\tau_{k-1}, \tau_k, \tau_{k+1})$. We know that the law of motion has the following property: conditional on $x(\tau_{k-1}) = 1 - x(\tau_k)$, and on $\tau_{k+1} - \tau_k = \tau_k - \tau_{k-1} = T$, it follows that, for any $t \in (0, T)$, $x(t + \tau_{k-1}) = 1 - x(t + \tau_k)$.

For τ_k large enough, $x(t+\tau_{k-1}) - x(t+\tau_k)$ is negligible because $E(\tau_{k+1} - \tau_k) = E(\tau_k - \tau_{k-1})$, the process is strongly mixing, and μ is large. Since matchings run independently of state-renewals, the distribution of the matching times on ω_1 approximates the distribution of the matching times on ω_2.

Since the sequence $\mathbf{x}$ on ω_2 depends only of $x(t+\tau_{k-1})$ and of the matching times on ω_2, and the sequence $1-\mathbf{x}$ on ω_1 depends only of $1-x(t+\tau_k)$ and of the matching times on ω_1, the two previous observations imply that $\Pr(\mathbf{x}|\omega_2) \approx \Pr(1-\mathbf{x}|\omega_1)$ for τ_k large enough, and μ large enough.

For $\lambda_1 \neq \lambda_2$, the above reasoning can be extended as long as $|\lambda_1 - \lambda_2| < M(\mu)$, for some $M(\mu)$, where as in Lemma 3, $M(\mu)$ is strictly positive, strictly increasing and $M(\mu) \to \infty$ for $\mu \to \infty$. ■

Proof of Proposition 5. Say without loss of generality that at matching m_k player j's history is $(a_2, 0)$. Player j must choose the optimal experimentation α out of the compact set $[0,1]$. Consider the coupling process previously introduced. Let $\mathcal{M}$ denote the set of all sequences $\mathbf{m}$ of matching times. Let $\Delta u(s) = 1$ for $s = (a_1, 1; a_2, 0), (a_1, 0; a_2, -1)$, $\Delta u(s) = -1$ for $s = (a_2, -1; a_1, 0), (a_2, 0; a_1, 1)$, and $\Delta u(C) = 0$.

The payoff difference at matching m_k is

$$\Delta U(a_2, 0, \varepsilon) = E\left[\Delta u(s) \Pr(s|t, \mathbf{m}, \omega, m_k, (a_2, 0))\right].$$

Since the population law of motion $x(t)$ is continuous in ε, looking at the matrix in the Proof of Lemma 2, one concludes $\Pr(s|t, \mathbf{m}, \omega, m_k, (a_2, 0))$ to be continuous in ε, so that, for all s, $\Delta u(s) \Pr(s|t, \mathbf{m}, \omega, m_k, (a_2, 0))$ is continuous in ε. Integrating a function continuous in ε against a measure independent of ε, we obtain a continuous $\Delta U(a_2, 0, \varepsilon)$.

Since $P(T_C < \infty) = 1$, $p(C, C) = 1$ and $\Delta U(a_2, 0, \varepsilon) = 0$ conditional on C, repeating the argument of the previous Lemma, the overtaking criterion implies that $\alpha = 1$ is optimal if and only if $\Delta U(a_2, 0, \varepsilon) > 0$, $\alpha = 0$ if and only if $\Delta U(a_2, 0, \varepsilon) < 0$, and, $\alpha = [0, 1]$ when $\Delta U(a_2, 0, \varepsilon) = 0$. Since $\Delta U((a_2, 0), 1) < 0$ (Lemma 4) and $\Delta U((a_2, 0), 0) > 0$ (Lemma 3), by continuity and the intermediate value theorem, there must exist a ε such that $\Delta U(a_2, 0, \varepsilon) = 0$. That implies $\varepsilon \in \alpha^\mu(\varepsilon) = [0, 1]$, so that such experimentation value ε is an equilibrium. In Lemma 4 we showed that for μ large, $\Delta U(a_2, 0, \varepsilon) < 0$ unless ε is small enough. That shows that $\lim_{\mu \to \infty} \underline{\varepsilon}(\mu) = 0$, where $\underline{\varepsilon}(\mu) = \sup\left\{\varepsilon \mid \varepsilon \in \alpha^\mu(\varepsilon)\right\}$. ■

Proof of Proposition 6. The first step consist of setting up the equilibrium condition on the auxiliary process.

Step 1: *Construction of the auxiliary process.*

Consider $T(\mu, \varepsilon)$, the amount of time needed for the population play $x|_2$ to reach $x(T) = 1/2$ starting from $x(0) = \bar{x}_\varepsilon$, when the state is ω_2 (by symmetry, that is also the lapse from $x(0) = \underline{x}_\varepsilon$ to $x(T) = 1/2$ and $\omega = 1$. Fix $\lambda = 1$, for μ large, the probability that a renewal occurs, before $x(T) = 1/2$ is reached, converges to 0. The population play when ω_1, will get arbitrarily close to $\bar{x}_\varepsilon$. Since $x'(\bar{x}_\varepsilon)$ is bounded away from 0, $T(\mu, \varepsilon)$ is close to the time

incurred between τ, the time an ω state switch and the time $T : x(T) = 1/2$, conditional on not there being any renewal between τ and T. As the expected time of renewal is $1/\lambda = 1$, by ergodicity $T(\mu, \varepsilon)$ may be considered also as the expected fraction of time spent by the system in the set $\{x(t) < 1/2, \omega_1\}$. Straightforward but tedious calculations from the law of motion yield:

$$T(\mu, \varepsilon) = \frac{\ln\left(\frac{1}{2\varepsilon\left(-2\varepsilon+\sqrt{(1+4\varepsilon^2)}\right)}\right)}{\mu\sqrt{(1+4\varepsilon^2)}}. \tag{5}$$

Note that $\mu T(\mu, \varepsilon) \to \infty$, for $\varepsilon \to 0$.

We write the indifference equation distinguishing between the fraction of time spent by the population with $x(t)$ □ $1/2$ and with $x(t) > 1/2$.

$$T(\mu, \varepsilon)\Delta U(\mu, \varepsilon, t \text{ □ } T) + (1 - T(\mu, \varepsilon))\Delta U(\mu, \varepsilon, t > T) = 0 \tag{6}$$

We know that $T(\mu, \varepsilon) \to \infty$ for $\varepsilon \to 0$ and $T(\mu, \varepsilon) \to 0$ for $\mu \to \infty$. Since we want to study the value of $T(\mu, \varepsilon)$ exactly for $\varepsilon \to 0$ and $\mu \to \infty$, we prefer to isolate the time T and rewrite Equation (6) as follows.

$$\begin{cases} 0 = T\Delta U(\mu, \varepsilon, t \text{ □ } T) + (1 - T)\Delta U(\mu, \varepsilon, t > T) \\ T = T(\mu, \varepsilon) \end{cases} \tag{7}$$

We will leave the second equation as calculated in Equation (5), and approximate the first equation for small ε and large μ. We shall obtain T as a function of μ and ε.

Consider the population law of motion on ω_1. Let by $x_{\mu,\varepsilon}(t, x_0)$ denote the solution of the related Cauchy problem with initial state x_0, and let $T_{\mu,\varepsilon}(x_0, x)$ be the solution of $x_{\mu,\varepsilon}(T, x_0) = x$. Fix small $\delta > 0$, straightforward calculations show that $\lim_{\varepsilon\to 0} \frac{T_{\mu,\varepsilon}(\underline{x}_\varepsilon, \underline{x}_\varepsilon+\delta)}{T_{\mu,\varepsilon}(\underline{x}_\varepsilon, 1/2)} = 1$, that $T_{\mu,\varepsilon}(1/2, \bar{x}_\varepsilon) = \infty$, and that $\lim_{\mu\to 0} T_{\mu,\varepsilon}(1/2, \bar{x}_\varepsilon - \delta) = 0$.

That means that, conditional on no state-switches occurring in the duration, for ε small and μ large, $\Pr(x(t) - \underline{x}_\varepsilon \in (0, \delta) | x(t) \text{ □ } 1/2) \approx 1$, and $\Pr(\bar{x}_\varepsilon - x(t) \in (0, \delta) | x(t) > 1/2) \approx 1$. As long as μ is large enough, the probability of a state switch is small enough, so that we can approximate $x(t)$ with $\underline{x}_\varepsilon + \delta/2$ on $t < T$, and with $\bar{x}_\varepsilon - \delta/2$ on $t > T$. Note moreover that $\lim_{\varepsilon\to 0} \lim_{\delta\to 0} \frac{\underline{x}_\varepsilon+\delta}{\varepsilon} = 1$ so we can approximate $x(t) \approx \varepsilon$ on t □ T and similarly, $x(t) \approx 1 - \varepsilon$ on $t > T$. After such simplifications, we obtain that $\Pr(\omega = 1 | (a_2, 0), t < T) = 1 - \varepsilon$ and $\Pr(\omega = 2 | (a_2, 0), t > T) = 1 - \varepsilon$.

Step 2: *The equilibrium experimentation $\varepsilon(\mu)$ is approximated by*

$$\left(\frac{1}{2}\mu T - \frac{1}{3}\mu^2 T^3\right)\varepsilon + \frac{1}{2}\mu T^2 - 1 \approx 0, \tag{8}$$

With the simplification obtained in Step 1 we can approximate the transitions in the coupling process by omitting all terms of order ε^2 and above. Proceeding in the same fashion as in Proof of Lemma 4, we derive that, when M meetings are yet to occur before time T is reached,

$$\Delta U(a_2, 0|\omega_1, M, t < T) = 1 + (1-\varepsilon)\sum_{t=0}^{M-2}(1-2\varepsilon)^t \approx 1 + (1-\varepsilon)\sum_{t=0}^{M-2}(1-2t\varepsilon) \approx M - (M+1)^2\,\varepsilon$$
$$\Delta U(a_2, 0|\omega_2, M, t > T) \approx -1 + (1-\varepsilon)\varepsilon(1-(1-2\varepsilon)) - \varepsilon(1-2\varepsilon) \approx -1 - \varepsilon$$

The distribution of M the remaining meetings in the reaction stage, given the time of meeting m_k, and the order of meeting k, is Poisson of parameter μT. When taking the expected value of a polynomial against a Poisson we obtain a higher ordered polynomial. As μT is large, we can eliminate all lower powered terms in μT and we can approximate $M \approx \mu T - k$. That is equivalent to treat the distribution of meetings as approximately uniform. By the same token, the distribution of the order of meeting k can be treated as an uniform draw out of $\mu T + 1$ outcomes (note that we are including the meeting of order 0, which is the last meeting before a state-renewal occurs).

$$\begin{aligned}\Delta U(a_2, 0|\omega_1, t \;<\; T) &= \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\Delta U(a_2, 0|\omega 1, M)\Pr(M|t<T, m_k)\Pr(k|t<T)\\ &\approx \sum_{k=0}^{\mu T}\frac{1}{\mu T+1}\Delta U(a_2, 0|\omega_1, \mu T - k) \approx \sum_{k=0}^{\mu T}\frac{\mu T - k - (\mu T - k + 1)^2\,\varepsilon}{\mu T + 1} =\\ &\quad \frac{1}{2}\mu T + \left(-\frac{1}{3}\mu^2 T^2 - \frac{7}{6}\mu T - 1\right)\varepsilon\end{aligned}$$

Now we can wrap up equation (7): after approximating it to eliminate higher powered ε terms and lower powered μT terms we obtain:

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

Step 3: *For μ large, $\varepsilon(\mu) \approx \frac{1}{2}e^{-\sqrt{2\mu}}$.*

Equation(8) is a cubic equation that has only one admissible solution $T(\mu, \varepsilon) \in (0, 1)$, the solution is continuous in ε. Since we are interested in approximation for small ε, we first solve for $T(\mu, \varepsilon)$ in Equation(8), with $\varepsilon = 0$. The equation $\frac{1}{2}\mu T^2 - 1 = 0$ has solution:

$$T = \frac{1}{\sqrt{\mu}}\sqrt{2}$$

Then we apply Dini's Theorem to obtain:

$$D_\varepsilon(T(\varepsilon,\mu)) = \frac{D_\varepsilon(\left(\frac{1}{2}\mu T - \frac{1}{3}\mu^2 T^3\right)\varepsilon + \frac{1}{2}\mu T^2 - 1)}{D_T(\left(\frac{1}{2}\mu T - \frac{1}{3}\mu^2 T^3\right)\varepsilon + \frac{1}{2}\mu T^2 - 1)} \approx -\frac{\frac{1}{2}\mu\frac{1}{\sqrt{\mu}}\sqrt{2} - \frac{1}{3}\mu^2(\frac{1}{\sqrt{\mu}}\sqrt{2})^3}{\mu\frac{1}{\sqrt{\mu}}\sqrt{2}} = \frac{1}{6}$$

So that, we can write a linear approximation for μ large and ε small:

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

We can finally compare this with the expression for T with $T(\varepsilon,\mu)$, the actual time spent by the system for $x(t) < 1/2$, as calculated from the law of motion in equation (5), so that the equilibrium condition can be approximated the following equation, which, comfortingly, displays $\varepsilon \to 0$, for $\mu \to \infty$.

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

After some further approximation, we obtain:

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

When $\lambda_1 \neq \lambda_2$, system (7) is modified so as to allow for two different times $T_1(\mu,\varepsilon)$ and $T_2(\mu,\varepsilon)$ associated to the states ω_1 and ω_2. For $\mu \to \infty$, it follows that $T_1(\mu,\varepsilon)/T_2(\mu,\varepsilon) \to 1$ from the law of motion. So that all the solutions of the system above will be of the same order in μ, and rest of the analysis is similar to the previous treatment. ■

References

[1] Banerjee A.V. [1992]: "A Simple Model of Herd Behavior" *The Quarterly Journal of Economics*, **107**: 797-817.

[2] Bjornestedt J. [1993]: "Experimentation, Imitation and Evolutionary Dynamics" *University of Stockolm*, mimeo

[3] Bjornestedt J. and J. Weibull [1993]: "Nash Equilibrium and Evolution by Imitation" in K. Arrow and E. Colombatto (eds.) *Rationality in Economics*, MacMillan New York

[4] Boylan R. [1992]: "Laws of Large Numbers for Dynamical Systems with Randomly Matched Individuals" *Journal of Economic Theory* 57: 473-504

[5] Borgers and Sarin [1999]: "Learning through Reinforcement and Replicator Dynamics" *Journal of Economic Theory* **77**: 1-14

[6] Cabrales and Sobel [1992]: "On the Limit Points of Discrete Selection Dynamics" *Journal of Economic Theory* **57**: 407-419

[7] Davis M. [1993]: *Markov Models and Optimization* Chapman and Hall, London

[8] Dekel and Scotchmer [1992]: "On the Evolution of Optimizing Behavior" *Journal of Economic Theory* **57**: 392-406

[9] Durrett R. [1996]: *Probability Theory and Examples* Belmont CA, Wadsworth Pu.

[10] Ellison and Fudenberg [1993]: "Rules of Thumb for Social Learning" *Journal of Political Economy* **51**: 612-43

[11] Friedman [1991]: "Evolutionary Games in Economics" *Econometrica* **59**: 637-666

[12] Hofbauer J. and J.W. Weibull [1996]: "Evolutionary Selection against Dominated Strategies" *Journal of Economic Theory* **71**: 558-73

[13] Keller G. and S. Rady [1999]: "Optimal Experimentation in a Changing Environment" *Review of Economic Studies* forthcoming

[14] Nachbar [1990]: "Evolutionary Selection Dynamics in Games: Convergence and Limit Properties" *International Journal of Game Theory* **19**: 59-89

[15] Moscarini G. Ottaviani M. and L. Smith [1998]: "Social Learning in a Changing World" *Economic Theory* **11**: 657-65.

[16] Rustichini A. and A. Wolinsky [1995]: "Learning about a Variable Demand in the Long Run" *Journal of economic Dynamics and Control* **19**: 1283-1292

[17] Smith L. and P. Sorensen [1999]: "Pathological Outcomes of Observational Learning" *Econometrica*, forthcoming

[18] Samuelson and Zhang [1992]: "Evolutionary Stability in Asymmetric Games" *Journal of Economic Theory* **57**: 363-391

[19] Schlag K. [1998]: "Why Imitate, and If So, How?" *Journal of Economic Theory* **78**: 130-156

[20] Weibull J. W. [1995]: *Evolutionary Game Theory* Cambridge MA, MIT press