JOURNAL OF ECONOMIC THEORY **74**, 235–265 (1997)
ARTICLE NO. ET962255

Muddling Through: Noisy Equilibrium Selection*

Ken Binmore

Department of Economics, University College London, London WC1E 6BT, England

and

Larry Samuelson

Department of Economics, University of Wisconsin, Madison, Wisconsin 53706

Received March 7, 1994; revised August 4, 1996

This paper examines an evolutionary model in which the primary source of "noise" that moves the model between equilibria is not arbitrarily improbable mutations but mistakes in learning. We model strategy selection as a birth–death process, allowing us to find a simple, closed-form solution for the stationary distribution. We examine equilibrium selection by considering the limiting case as the population gets large, eliminating aggregate noise. Conditions are established under which the risk-dominant equilibrium in a 2×2 game is selected by the model as well as conditions under which the payoff-dominant equilibrium is selected. *Journal of Economic Literature* Classification Numbers C70, C72.

Commonsense is a method of arriving at workable conclusions from false premises by nonsensical reasoning.

Schumpeter

1. INTRODUCTION

Which equilibrium should be selected in a game with multiple equilibria? This paper pursues an evolutionary approach to equilibrium selection based on a model of the dynamic process by which players adjust their strategies.

* First draft August 26, 1992. Financial support from the National Science Foundation, Grants SES-9122176 and SBR-9320678, the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303, and the British Economic and Social Research Council, Grant L122-25-1003, is gratefully acknowledged. Part of this work was done while the authors were visiting the University of Bonn, for whose hospitality we are grateful. We thank Drew Fudenberg, George Mailath, Reinhard Selten, Avner Shaked, Richard Vaughan, an associate editor, and two anonymous referees for helpful discussions and comments.

0022-0531/97 $25.00

A more orthodox approach to the equilibrium selection problem is to invent refinements of the Nash equilibrium concept. In the same spirit, numerous refinements of the notion of an evolutionarily stable strategy have been proposed. From this perspective, it may be troubling that the equilibrium selected by a dynamic model such as ours often depends on the fine details of the modeling or on the initial conditions prevailing at the time the process began. But we consider this dependence to be a virtue rather than a vice. The institutional environment in which a game is learned and played can matter for equilibrium selection. Theories of equilibrium selection cannot neglect such details. Instead, we must be explicit about which aspects of a game's environment and the process by which players learn to play the game are significant and how they determine which equilibrium is selected.

In Binmore *et al.* [7] we discussed the differences between the long-run and ultralong-run behavior of an evolutionary model. Our concern in this paper is with equilibrium selection in the ultralong run. The "ultralong run" refers to a period of time sufficiently long, not only for trial-and-error learning to direct the agents in our model to an equilibrium, but also for random shocks to bounce the system repeatedly from one equilibrium into the basin of attraction of another, so establishing a steady-state frequency with which each equilibrium is visited. If all but one of the equilibria are visited with negligible frequency, then we say that the remaining equilibrium is "selected" in the ultralong run.[1]

The pioneers in extracting ultralong-run equilibrium selection results from explicit learning models are Kandori *et al.* [19] and Young [33]. In the Kandori, Mailath and Rob model, agents choose best responses given their information, prompting us to describe them as *maximizers*. However, after agents have decided on an action, there is a small probability $\lambda > 0$ that they will switch their choice to some suboptimal alternative. Such switches are said to be *mutations*. The ultralong-run distribution over population states is then studied in the limit as $\lambda \to 0$. A striking prediction of both Kandori *et al.*'s and Young's models is that, as this limit is taken, the distribution over population states concentrates all of its probability on the risk-dominant equilibrium in 2×2 symmetric games.

This paper is motivated by a simple belief: that people make mistakes. It may be that people are more likely to switch to a best reply than otherwise, but people are unlikely to be so flawless that they *always* switch to a best reply when reassessing their strategies. Furthermore, we do not expect

[1] Kandori *et al.* [19] call such an equilibrium a "long-run equilibrium." We reserve the term "long run" for a period of time long enough for the system to reach the first equilibrium it will visit. We consider long-run phenomena important but concentrate on the ultralong run in this paper.

these mistakes to be negligible, and hence do not think it appropriate to examine the limiting case as the mistakes become arbitrarily small. We refer to agents who are plagued by such mistakes as *muddlers* and refer to such learning as *noisy* learning. These mistakes might seem egregious in the stark models with which we commonly work, but arise quite naturally in the noisy world in which games are actually played.

Kandori *et al.* [19] and Young [13] obtain tractable and precise results from their models by considering the limit as noise, in the form of mutations, becomes vanishingly small. Given our belief that mistakes in learning are not insignificant, we cannot adopt a similar approach. Instead, we examine a model in which at most one agent at a time changes strategies. This yields a birth–death process for which we can obtain a simple closed-form solution even when the noise is bounded away from zero.[2] We pay a price for this convenience: the stationary distribution typically does not concentrate its probability around a single equilibrium. Instead, the nonnegligible noise ensures that the distribution disperses a significant fraction of its probability throughout the state space. We show, however, that as the size of the population grows, *aggregate* noise is eliminated and the stationary distribution in 2×2 symmetric games becomes increasingly concentrated near a single equilibrium, allowing equilibrium selection results to be obtained for large populations.

Noisy learning has implications for equilibrium selection. In the symmetric 2×2 games studied in this paper, maximizing models choose between two strict Nash equilibria by selecting the risk-dominant equilibrium. When risk-dominance and payoff-dominance conflict, our muddling model sometimes selects the payoff-dominant equilibrium.

Examining muddlers rather than maximizers also has the consequence that the expected waiting time before the ultralong-run predictions of the model become relevant can be greatly reduced. To see why, consider the possibility that a population of agents has found its way to an equilibrium that is not selected in the ultralong run. In the maximizing models of Kandori *et al.* [19] and Young [33], a large number of effectively *simultaneous* mutations are now necessary for the system to escape from the equilibrium's basin of attraction. In contrast, our muddling model requires only one mutation to step away from the equilibrium, after which the agents may *muddle* their way out of its basin of attraction. If the mutation probability is small, multiple mutations are much less likely than a single mutation and the former model accordingly has much longer expected waiting time. At the same time, we again note that our model yields precise equilibrium selection results only in the limit as the population size gets large, and our

[2] Amir and Berninghaus [1] provide a complementary analysis based on a birth–death process.

expected waiting times grow as does the population. Our relatively rapid convergence thus comes at the cost of noisier equilibrium selection results.

Section 2 presents the muddling model. Section 3 examines the dynamics of the resulting equations of motion and takes up the problem of expected waiting times. Section 4 discusses ultralong-run equilibrium selection for the muddling model. The results of Sections 2–4 depend only on the assumptions that there is some tendency for agents to move in the direction of a best reply and that they occasionally make mistakes in doing so.

Section 5 turns to a specific example of the learning process. In the context of this example, we derive conditions under which the payoff-dominant or risk-dominant equilibrium will be selected. Section 6, with the help of yet more structure, considers the evolutionary stability of the learning rules studied. We ask whether a population using a certain learning rule, and hence receiving the payoffs associated with the corresponding ultralong-run distribution over population states, can be invaded by a mutant learning rule from the same (narrow) class of learning rules. If it can, then we have grounds for questioning its robustness. We find conditions under which evolution has a tendency to select learning rules that in turn yield the risk-dominant equilibrium for symmetric 2×2 games.

2. A MUDDLING MODEL

The Game. We begin with the symmetric 2×2 game $\mathcal{G}$ of Fig. 1. We assume that there is a single population containing N agents. Time is divided into discrete intervals of length τ. In each time period, an agent is characterized by the strategy X or Y that she is programmed to use in that period. An agent playing X receives a payoff of A in a population in which all agents play X. We assume that the process by which agents are matched to play the game is such that an agent playing X receives an expected payoff of $\pi_X(k) = kA + (1-k)\,C$ when proportion k of her opponents play X and proportion $(1-k)$ play Y. She receives $\pi_Y(k) = kB + (1-k)\,D$ when playing Y under similar circumstances.

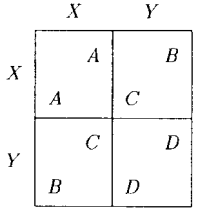

FIG. 1. The game $\mathcal{G}$.

Muddled Learning. We consider a general model of muddled learning and a specific example in which much sharper assumptions are made. The general model is built around Assumptions 1–3.

The model has four parameters: the time t at which the system is observed, the length τ of a time period, the population size N, and the mutation rate λ. The ultralong-run behavior of the system is examined by taking the limit $t \to \infty$. We then take the limit $\tau \to 0$. This gives a model in which agents revise their strategies at uncoordinated, idiosyncratic times.[3] Finally, we take the limits $N \to \infty$ and $\lambda \to 0$ in order to sharpen the results. A discussion of the implications of taking limits in different orders appears in Binmore *et al.* [7].

A population state x is the number of agents currently playing strategy X. The fraction of such agents is denoted by $k = x/N$. Learning is taken to be an infrequent occurrence compared with the play of the game. At the end of each period of length τ, a mental bell rings inside each player's head, where the units in which time is measured are chosen so that the probability of such an occurence is τ. An agent for whom the bell tolls is said to receive the learn-draw.

Learn-draws are independent across agents and across time. An agent who does not receive the learn-draw retains her current strategy while an agent receiving the learn-draw potentially changes strategies.

Because we consider the case $\tau \to 0$, occurrences in which more than one agent receives the learn-draw in a single period will be very rare. As a result, we will find that the system can be described in terms of the probabilities that, when a learn-draw is received, the number of agents currently playing strategy X increases or decreases by one. Let $r_{(\lambda, N)}(x)$ be the probability that, given a single player (and only a single player) receives the learn-draw, the result is to cause a player currently strategy Y to switch to X. (Hence $r_{(\lambda, N)}(N) = 0$.) Similarly, let $\ell_{(\lambda, N)}(x)$ be the probability that, given a single player receives the learn-draw, the result is to cause a player currently playing X to switch to Y. (Hence $\ell_{(\lambda, N)}(0) = 0$.)

We think of the parameter $\lambda \geqslant 0$ that appears in $r_{(\lambda, N)}(x)$ and $\ell_{(\lambda, N)}(x)$ as the rate of mutation, where "mutation" is a catch-all term for a variety of minor disturbances that modelers would normally suppress in the belief that they are too small to be relevant. Since our focus will be on what happens as $\lambda \to 0$, we assume that $r_{(\lambda, N)}(x)$ and $\ell_{(\lambda, N)}(x)$ are continuous on the right at $\lambda = 0$. We refer to $r_{(0, N)}(x)$ and $\ell_{(0, N)}(x)$ as the *learning process.*

[3] Other approaches will be necessary if strategy revisions are coordinated, perhaps by the regular arrival of information in an economic context or by the presence of breeding seasons in a biological context.

Assumption 1.

[1.1] $r_{(0,N)}(0) = \ell_{(0,N)}(N) = 0$;

[1.2] $(x \in \{1, ..., N-1\}, \lambda > 0) \Rightarrow r_{(\lambda,N)}(x) > h_N(x) > 0$;

[1.3] $(x \in \{1, ..., N-1\}, \lambda > 0) \Rightarrow \ell_{(\lambda,N)}(x) > h_N(x) > 0$;

[1.4] $(x \in \{1, ..., N-1\}, \lambda > 0) \Rightarrow 0 < h < r_{(\lambda,N)}(x)/\ell_{(\lambda,N)}(x) \leqslant H < \infty$;

[1.5] $\lim_{\lambda \to 0} r_{(\lambda,N)}(0)/\lambda \in (h, H)$; $\lim_{\lambda \to 0} \ell_{(\lambda,N)}(N)/\lambda \in (h, H)$;

where $h_N(x)$ is a positive-valued function on $\{1, ..., N-1\}$ and h and H are constants. Assumption 1.1 asserts that the learning process alone cannot cause an agent to switch to a strategy not already present in the population. This assumption is not strictly necessary but we consider it realistic.[4] Assumptions 1.2–1.3 are the essence of the muddling model. They ensure that $r_{(0,N)}(x)$ and $\ell_{(0,N)}(x)$ are positive (except in the pure population states $x = 0$ and $x = N$). The learning process may thus either increase or decrease the number of agents playing strategy X, and hence may either move agents toward or away from best replies, as long as both strategies are present in the population. We shall shortly interpret Assumption 1.4 as ensuring that our muddling agents are not too close to being maximizers. It is important here that h and H do not depend on λ or N. Assumption 1.5 states that mutations can push the population away from a state in which all agents play the same strategy and ensures that the probability of switching away from a monomorphic state is of the same order of magnitude (for small mutation probabilities) as the probability of a mutation. We interpret this as the statement that a single mutation suffices to introduce a new strategy into a monomorphic population.

Assumption 2. There exist functions $r_\lambda(k)$ and $\ell_\lambda(k)$ which are continuous in λ and k for $0 \leqslant \lambda \leqslant 1$ and $0 \leqslant k \leqslant 1$ such that

[2.1] $r_{(\lambda,N)}(kN) = r_\lambda(k) + O(1/N)$

[2.2] $\ell_{(\lambda,N)}(kN) = \ell_\lambda(k) + O(1/N)$.

Assumption 2 is the requirement that only the *fraction* of agents playing X is significant when N is large.

Assumption 3.

[3.1] $0 < k < 1 \Rightarrow (r_0(k) > 0$ and $\ell_0(k) > 0)$;

[3.2] $\pi_X(k) > \pi_Y(k) \Leftrightarrow r_0(k) > \ell_0(k)$;

[3.3] $\pi_X(k) < \pi_Y(k) \Leftrightarrow r_0(k) < \ell_0(k)$.

[4] It is natural when learning is driven by imitation or when changes in the composition of the population are caused by biological reproduction.

Assumption 3.1 ensures that the muddling present in the learning process does not disappear as the population grows large. Assumptions 3.2–3.3 require that the learning process is always more likely to move in the direction of a best reply than away from it. In light of this, Assumption 1.4 has the effect of preventing the probability of moving in the direction of the best reply from becoming arbitrarily large compared with the alternative and hence ensures that our muddling agents are not arbitrarily close to being maximizers.[5]

Aspirations and Imitation: An Example. This section presents a simple learning model satisfying Assumptions 1–3. Binmore *et al.* [7] present a biological example.

In each period of length τ, pairs of agents are randomly drawn (independently and with replacement) to play the game. Such draws occur sufficiently frequently that the probability of each agent playing at least one game in each period can be taken to be unity. Given that agents are drawn randomly and with replacement, this implies that each agent will have played an infinite number of games with a distribution of opponents that accurately reflects the distribution of strategies in the population.[6]

For this example, we interpret the payoffs given in game $\mathcal{G}$ as *expected payoffs*. Realized payoffs are random, being given by the average expected payoff in the game $\mathcal{G}$ plus the outcome R of a random variable $\tilde{R}$ which potentially captures a variety of random shocks that perturb payoffs.[7] This randomness in turn may be an important reason why learning proceeds in a muddling rather than maximizing fashion.

[5] Blume [8] examines a model satisfying Assumption 3, with agents being more likely (but not certain) to switch to high-payoff strategies and with switching probabilities being smoothly related to payoffs.

[6] Since we consider the case when $\tau \to 0$, this assumption is a very strong version of requiring that the game be played arbitrarily rapidly. We view this as an approximation of the case when play is frequent relative to strategy revision, which we consider a natural setting for evolutionary models. Kandori *et al.* [19] assume that agents play an infinite number of times in each period or that a round-robin tournament is played in each period. Nöldeke and Samuelson [25] assume a round-robin tournament. Young's model [33] is less demanding in this respect, though all agents still have access to the result of each game as soon as it is played. An alternative model which assumes that agents do *not* play all other agents and which exploits this fact to obtain short waiting times, is examined by Robson and Vega Redondo [27].

[7] The random variable $\tilde{R}$ yields a shock common to *each* payoff received by an agent in the given period. The distribution F of $\tilde{R}$ is independent and identically distributed across players. It would also be interesting to study cases in which the distribution of $\tilde{R}$ differs across players, or in which this source of noise is correlated across individuals, perhaps as a result of environmental factors that impose a common risk on all agents. Papers in which the latter type of uncertainty appears include Fudenberg and Harris [14] and Robson [26].

An agent who receives the learn-draw recalls her average realized payoff in the last period and assesses it as being either "satisfactory" or "unsatisfactory."[8] If the average realized payoff exceeds an aspiration level, then the strategy is deemed satisfactory and the agent makes no change in strategy. If instead the average realized payoff falls short of the aspiration level, then the agent loses faith in her current strategy and abandons it. We refer to the probabilities that a player who has received the learn-draw will abandon her current strategy as *death probabilities* in order to stress the mathematical parallels between this model and that of Binmore *et al.* [7]. For each expected payoff π, the corresponding death probability is given by

$$g(\pi) = \text{prob}(\pi + R < \Delta) = F(\Delta - \pi), \tag{1}$$

where Δ is the aspiration level, R is the realization of the random payoff variable $\tilde{R}$, and F is the cumulative distribution of $\tilde{R}$.

We assume that F is log-concave.[9] A sufficient condition for the log-concavity of F is that its density f be log-concave (Bagnoli and Bergstrom [2, Lemma 1]), or equivalently that f satisfy the monotone likelihood ratio property. The latter is a necessary and sufficient condition for it to be more likely that low average realized payoffs are produced by low average expected payoffs, and hence for realized payoffs to provide a useful basis for evaluating strategies (see Milgrom [23]).

If agent i has abandoned her strategy as unsatisfactory, she must now choose a new strategy. We assume that she randomly selects a member j of the population. With probability $1-\lambda$, i imitates j's strategy.[10] With probability λ, i is a "mutant" who chooses the strategy that j is not playing.

We refer to this as the *aspiration and imitation* model. The fact that we are free to specify the aspiration level Δ and the distribution F allows several familiar formulations to appear as special cases.[11] For example, suppose that the payoffs satisfy $A > D > B > C$, so that the game has two

[8] Satisficing models have long been the primary alternative to models of fully rational behavior, being pioneered in economics by Simon [28–30] and in psychology by Bush and Mosteller [9], and pursued in such work as Winter [32] and Nelson and Winter [24]. More recently, satisficing models built on aspiration levels have been examined by Bendor *et al.* [3] and Gilboa and Schmeidler [15–17].

[9] This means that $\ln F$ is concave or, equivalently, that f/F is decreasing, where f is the density of the cumulative distribution F. See Bagnoli and Bergstrom [2] for a discussion of log-concavity and its implications. Many common distributions are log-concave, including the chi, chi-squared, exponential, logistic, normal, Pareto, Poisson, and uniform distributions.

[10] She may thereby end up playing the strategy with which she began, having perhaps had her faith in it restored by seeing it played by the person she chose to copy.

[11] Even more flexibility could be obtained by allowing the aspiration level to differ across agents and across states, perhaps depending upon prevailing payoffs. This is consistent with our general model, but we do not pursue it here in order to keep the example simple.

strict Nash equilibria. If we choose F to put a probability mass of one on the value zero and take Δ to be the payoff of the mixed strategy equilibrium of the game, then we have random-best-reply dynamics, with agents who are chosen to learn switching strategies only if their current strategy is not a best reply.[12]

An interesting special case is that in which F is the uniform distribution on the interval $[-\omega, \omega]$, where $\{A, B, C, D\} \subset [\Delta - \omega, \Delta + \omega]$. Death probabilities are then linear in expected payoffs. In this case, the model is equivalent to one in which each agent plays only once in each period.

We can calculate $r_{(\lambda, N)}(x)$ for the aspiration and imitation model. For the number of agents playing X to increase, given that an agent has received the learn-draw, three events must occur: (i) The agent who receives the learn-draw must be playing strategy Y. If x agents are currently playing strategy X, then the probability that an agent drawn to learn is playing strategy Y is given by $(N-x)/N$. (ii) The learning agent must abandon her current strategy. Because the average payoff of an agent playing strategy Y is $(xB+(N-x-1)\,D)/(N-1)$, this occurs with probability $g((xB+(N-x-1)\,D)/(N-1))$, where g is defined by (1). (iii) The learning agent must choose X for her new strategy. This occurs with probability $((1-\lambda)\,x+\lambda(N-x-1))/(N-1)$, since with probability $(1-\lambda)\,x/(N-1)$, the learning agent chooses to imitate an agent playing X and does so without mutation, and with probability $\lambda(N-x-1)/(N-1)$ the learning agent chooses to imitate an agent playing Y but is a mutant and chooses strategy X. Putting these probabilities together, we have

$$r_{(\lambda, N)}(x) = \frac{N-x}{N}\, g\left(\frac{xB+(N-x-1)\,D}{N-1}\right)\frac{(1-\lambda)\,x+\lambda(N-x-1)}{N-1}. \quad (2)$$

Similarly,

$$\ell_{(\lambda, N)}(x) = \frac{x}{N}\, g\left(\frac{(x-1)\,A+(N-x)\,C}{N-1}\right)\frac{\lambda(x-1)+(1-\lambda)(N-x)}{N-1}. \quad (3)$$

Combining these for the case where $\lambda \to 0$ and $N \to \infty$, we have

$$\frac{r_0(k)}{\ell_0(k)} = \frac{g(\pi_Y(k))}{g(\pi_X(k))}. \quad (4)$$

[12] It may appear counterintuitive to speak of best-reply dynamics when agents are choosing strategies by simply imitating others, but a model in which agents abandon only inferior replies but choose strategies by imitation is analogous to a model in which agents are randomly chosen to switch to best replies.

3. DYNAMICS

Stationary Distribution. To examine the ultralong-run behavior of our learning model, we study the stationary distribution of the system. For a fixed set of values of the parameters τ, λ, and N, we have a homogeneous Markov process $\Gamma_{(\lambda, N, \tau)}$ on a finite state space. In addition, the Markov process is irreducible, because Assumptions 1.2, 1.3 and 1.5 ensure that for any state $x \in \{0, 1, ..., N\}$, there is a positive probability both that the Markov process moves to the state $x+1$ (if $x < N$), in which the number of agents playing X is increased by one; and that the process moves to the state $x-1$ (if $x > 0$), in which the number of agents playing X is decreased by one. The following result is then standard:

PROPOSITION 1. *The Markov process $\Gamma_{(\lambda, N, \tau)}$ has a unique stationary distribution. For any initial condition, the expected proportion of time to date T spent in each state converges as $T \to \infty$ to the corresponding stationary probability; and the distribution over states at a given time T converges to the stationary distribution.*

Proof. Kemeny and Snell [21, Theorems 4.1.4, 4.1.6, and 4.2.1.]. ∎

Let $\gamma_{(\lambda, N, \tau)}$ be the probability measure induced by the stationary distribution, with $\gamma_{(\lambda, N, \tau)}$ hereafter simply called the stationary distribution. Then $\gamma_{(\lambda, N, \tau)}(x)$ is the probability attached to state x. We study the distribution $\gamma_{(\lambda, N)}$ obtained from $\gamma_{(\lambda, N, \tau)}$ by taking the limit $\tau \to 0$.

As $\tau \to 0$, the event in which more than one agent receives the learn-draw occurs with negligible probability. The model is thus a birth–death process, as studied in Karlin and Taylor [20, Chap. 4]. The following result is standard, where (5) is known as the "detailed balance" equation:[13]

[13] See, for example, Karlin and Taylor [20, p. 137]. The techniques of Freidlin and Wentzell have become common, and can be used to give an alternative proof of this result. Freidlin and Wentzell [12, Lemma 3.1 on p. 177] show that $\gamma_{(\lambda, N)}(x+1)/\gamma_{(\lambda, N)}(x)$ is given by the ratio of the sum of the products of the probabilities of the transitions in all $(x+1)$-trees to the similar calculation for x-trees (where an $(x+1)$-tree is a collection of transitions with the properties that every state other than $x+1$ is the origin of one and only one transition, there is a path of transitions from every state except $x+1$ to $x+1$, and there are no cycles). In the limit as τ becomes small, the only trees that are relevant are those that involve no transitions that occur with probability τ^2 or less, i.e., involve only transitions from a state to one of its immediate neighbors. There is only one such tree for each of states $x+1$ and x, consisting of a transition from each state other than $x+1$ (or x) to the immediate neighbor that lies closest to $x+1$ (or x). These two trees differ only in one probability: The $x+1$-tree contains the probability $r_{(\lambda, N)}(x)$ while the x-tree contains $\ell_{(\lambda, N)}(x+1)$, giving (5).

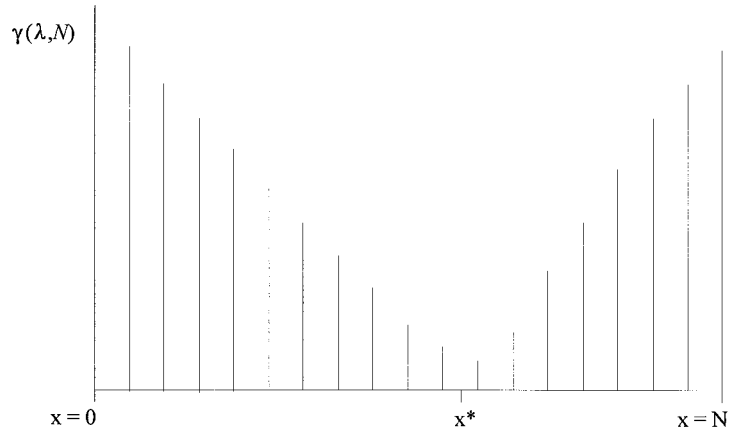

FIG. 2. Stationary distribution.

PROPOSITION 2. *Consider states x and $x+1$. Then the limiting stationary distribution* $\lim_{\tau \to 0} \gamma_{(\lambda, N, \tau)} = \gamma_{(\lambda, N)}$ *exists and satisfies*:

$$\frac{\gamma_{(\lambda, N)}(x+1)}{\gamma_{(\lambda, N)}(x)} = \frac{r_{(\lambda, N)}(x)}{\ell_{(\lambda, N)}(x+1)}. \tag{5}$$

To interpret (5), consider a game with two strict Nash equilibria. Let k^* be the probability attached to X by the mixed-strategy Nash equilibrium of the game and let $x^*/N \equiv k^*$. (Note that x^* need not be an integer.) Then if $x > x^*$, we must have $\gamma_{(\lambda, N)}(x+1) > \gamma_{(\lambda, N)}(x)$ if λ is sufficiently small and N large, (because strategy X must be a best reply here, so that Assumption 3.2 gives $r_0(x/N) > \ell_0(x/N)$, and then Assumptions 2 and 3.1 give $r_{(\lambda, N)}(x) > \ell_{(\lambda, N)}(x+1)$). The stationary distribution $\gamma_{(\lambda, N)}$ must then increase on $[x^*, N]$. Similarly, from Assumption 3.3, $\gamma_{(\lambda, N)}(x+1) < \gamma_{(\lambda, N)}(x)$, and $\gamma_{(\lambda, N)}(x)$ must decrease on $[0, x^*]$.[14] The graph of γ therefore reaches maxima near the endpoints of the state space. These endpoints correspond to the strict Nash equilibria of the game at which either all agents play X or all agents play Y. Its minimum is achieved at x^*, as shown in Fig. 2.

The convenience of the detailed balance Eq. (5) is available because we have restricted attention to two-player 2×2 games. Extending the analysis to larger games will require new techniques, though some of the basic ideas will reappear. In particular, we hope to establish conditions under which the stationary distribution concentrates its probability near equilibria of

[14] The precise statement here is that for fixed $\varepsilon > 0$, there is sufficiently large N and small λ such that $\gamma_{(\lambda, N)}$ increases on $[x^* + \varepsilon, N - \varepsilon]$, decreases on $[\varepsilon, x^* - \varepsilon]$, and whose minimum and maximum on $[x^* - \varepsilon, x^* + \varepsilon]$ differ by less than ε.

larger games (cf. Proposition 4 below). Ascertaining how the model selects between these equilibria will then require calculations involving the relative likelihoods of transitions between equilibria. In a two-player 2×2 game, there is only one way such a transition can proceed, and calculating the relative likelihoods of the transitions is straightforward (cf. Corollary 1 below). In larger games, there will be many routes by which such a transition might proceed, and considerable work remains to be done in transforming these multiple routes into tractable conditions.

Convergence. How long is the ultralong run? We provide a comparison of the convergence properties of our muddling model and the model of Kandori *et al.* [19]. To do so, we fix the population size N and follow Kandori *et al.* in examining the limit as the probability of a mutation λ gets small. We consider the case of a game with two strict Nash equilibria.

How is our continuous-time model to be compared to the discrete model of Kandori *et al.*? Fix the unit in which time is to be measured. This unit of measurement will remain constant throughout the analysis, even as we allow the length of the time periods between learn-draws in our muddling model to shrink. Our question then concerns how much time, measured in terms of the fixed unit, must pass before the probability measure describing the expected state of the relevant dynamic process is sufficiently close to its stationary distribution. To make the models comparable, we assume that the episodes in which every agent learns in the Kandori *et al.* model occur at each of the discrete times 1, 2, …. We then recall that τ is the probability of a birth per τ units of time in our model. In the limit as $\tau \to 0$, the expected number of times in an interval of time of length one (which will contain many of our very short time periods) that an agent in our model learns is then one, matching the Kandori *et al.* model.

Recall that $\Gamma_{(\lambda, N, \tau)}$ is the transition matrix for the Markov process of our muddling model given mutation rate λ and period length τ. The matrix $\Gamma_{(\lambda, N, \tau)}$ depends on τ because the probability of an agent receiving the learn-draw in a given period depends on the period length. Notice also that as τ decreases, the number t/τ of periods up to time t increases.

PROPOSITION 3. *There exists a function $h_\gamma(\lambda)$ such that for any initial distribution γ^0 and sufficiently small λ,*[15]

$$\lim_{\tau \to 0} \left(1\text{-}\limsup_{t \to \infty} \|\gamma^0[\Gamma_{(\lambda, N, \tau)}]^{t/\tau} - \gamma_{(\lambda, N)}\|^{1/(t-1)}\right) \leqslant h_\gamma(\lambda) \sim \lambda, \tag{6}$$

where γ^0 is the initial distribution.

[15] We say that the fusions $f(\lambda)$ and $g(\lambda)$ are comparable and write $f \sim g$, if there exist constants c and C such that for all sufficiently small λ, $c\,|g(\lambda)| \leqslant |f(\lambda)| \leqslant C\,|g(\lambda)|$. $\|\cdot\|$ is the max norm.

The gap between the distribution of the muddling Markov process at time t and the stationary distribution, given by $\|\gamma^0[\Gamma_{(\lambda, N, \tau)}]^{t/\tau} - \gamma_{(\lambda, N)}\|$, thus decreases exponentially, with the gap eventually decreasing to at most $(1-h_\gamma(\lambda))^{t-1}$. It is a standard result that finite Markov processes converge at exponential rates. The proof, contained in the Appendix, involves a straightforward calculation of the rate and an examination of the limit $\tau \to 0$.

Let $\Psi_{(\lambda, N)}$ be the transition matrix of the Kandori *et al.* model given mutation rate λ and population size N, and let $\psi_{(\lambda, N)}$ be its stationary distribution. Ellison [10] shows that there exists a function $h_\psi: \mathbb{R} \to \mathbb{R}$ such that

$$1\text{-}\sup_{\psi^0} \limsup_{t \to \infty} \|\psi^0[\Psi_{(\lambda, N)}]^t - \psi_{(\lambda, N)}\|^{1/t} = h_\psi(\lambda^z) \sim \lambda^z, \tag{7}$$

where ψ^0 is the initial distribution, $\psi^0[\Psi_{(\lambda, N)}]^t$ is the distribution at time t, and z is the minimum number of an agent's opponents that must play the risk-dominant equilibrium strategy in order for the latter to be a best reply for the agent in question. Hence, we again have exponential convergence, at a rate that is arbitrarily close to $1-h_\psi(\lambda^z)$ for large t.

Together, (7) and (6) imply that for very small values of λ, the muddling model converges much faster than does the Kandori *et al.* model. In particular, let $T_\psi(\eta)$ be the length of time required for Kandori *et al.* model to be within η of its stationary distribution, i.e., for $\|\psi^0[\Psi_{(\lambda, N)}]^t - \psi_{(\lambda, N)}\| \leqslant \eta$. Then from (7), for sufficiently large t and hence sufficiently small η, we have the approximation $\eta = (1-h_\psi(\lambda^z))^{T_\psi(\eta)}$. Fixing a sufficiently large such $T_\psi(\eta)$ and small η, let $T_\gamma(\eta)$ be an analogous measure for our muddling model. Then from (7), we have the approximation $\eta \leqslant (1-h_\gamma(\lambda))^{T_\gamma(\eta)-1}$. These approximations can be made arbitrarily precise by taking t large, giving, for small values of λ and large t,

$$\frac{T_\psi(\eta)}{T_\gamma(\eta)-1} \geqslant \frac{\ln(1-h_\gamma(\lambda))}{\ln(1-h_\psi(\lambda^z))} \approx \frac{h_\gamma(\lambda)}{h_\psi(\lambda^z)} \sim \frac{1}{\lambda^{z-1}}. \tag{8}$$

If, for example, $N = 100$ and $z = 33$, so that 1/3 of one's opponents must play the risk-dominant strategy in order for it to be a best reply, then it will take $1/\lambda^{32}$ times as long for the Kandori *et al.* model to be within η of its stationary distribution as it takes the muddling model. Ellison [10] obtains a similar comparison for the Kandori, Mailath and Rob model and his "two-neighbor" matching model. Ellison notes that if $N = 100$ and $z = 33$, then halving the mutation rate causes his two-neighbor matching model (and hence our muddling model) to take about twice as long to converge, while the Kandori, Mailath and Rob model will take 2^{23} (> 8 billion) times as long to converge.

What drives this difference in rates of convergence? The Kandori *et al.* model relies upon mutations to accomplish its transitions between equilibria. For example, the stationary distribution may put all of its probability on state 0, but the initial condition may lie in the basin of attraction of state N. Best-reply learning then takes the system immediately to state N, and convergence requires waiting until the burst of z simultaneous mutations, required to jump over the basin of attraction of N and reach the basin of attraction of 0, becomes a reasonably likely event. Since the probability of such an event is of the order of λ^z, this requires waiting a very long time when the mutation rate is small. In contrast, the muddling model requires mutations only to escape boundary states (see Assumption 1). Once a single mutation has allowed this escape (cf. Assumption 1.5), then the noisy learning dynamics can allow the system to "swim upstream" out of its basin of attraction.[16] The probability of moving from state N to state 0 is given by $\prod_{x=N}^{1} \ell_{(\lambda, N)}(x)$. When mutation rates are small, the learning dynamics proceed at a much faster rate than mutations occur, so that only one term in this expression ($\ell_{(\lambda, N)}(N)$) is of order λ. Convergence then requires waiting only for a single mutation, rather than z simultaneous mutations, and hence relative waiting times differ by a factor of λ^{z-1}.

The difference in rates of convergence for these two models will be most striking when the mutation rate is very small. The comparison we have chosen, especially the examination of the limit as the mutation rate approaches zero, puts the Kandori *et al.* model in the worst possible case. In [7], we present an example in which $N = 100$, $z = 33$, and $\lambda = 0.001$. The expected waiting time in the Kandori *et al.* model is approximately 1.7×10^{72}, while that of the muddling model is approximately 5000. We expect the waiting times to be closer for larger mutation rates because increasing λ makes the Kandori *et al.* model noisier (and conceptually closer to a muddling model), reducing its waiting time. We would also expect the waiting times to be closer if we forced the noise in our learning process to become negligible. Both observations are variations on the point we want to make with the comparison. Incorporating nonnegligible noise into a model can hasten its convergence. Even if unexplained, exogenously determined perturbations (mutations) are to be treated as negligible, expected waiting times can still be short if one is realistic in building noise into the learning process itself.

We have examined waiting times for a fixed population size N. Our model yields sharp equilibrium selection results as $N \to \infty$, but need not do so for small values of N. In addition, the expected waiting time diverges in

[16] A similar distinction, including the "swimming upstream" analogy, appears in Fudenberg and Harris [14].

our model as N increases, becoming arbitrarily long as N gets large.[17] There accordingly remains plenty of room for skepticism as to the applicability of ultralong-run analyses based on examining stationary distributions. In particular, we can achieve crisp equilibrium selection results only at the cost of long waiting times, and the finding that waiting times can be short must be tempered by the realization that the resulting stationary distribution may give noisy equilibrium selection results. In many cases, however, a population that is not arbitrarily large and a stationary distribution that allocates probability to more than one state may be the most appropriate model, even though it does not give unambiguous equilibrium selection results.[18] Equilibrium selection is then not our only concern. Our model allows tractable, closed-form solutions to be obtained when the population is not so large as to eliminate aggregate noise from the model.

4. EQUILIBRIUM SELECTION

We now consider equilibrium selection. We concentrate on the case of large populations and small mutation rates. In particular, we begin with the limiting stationary distribution of the Markov process as $\tau \to 0$ and then study the limits $N \to \infty$ and $\lambda \to 0$. The order in which these two limits are taken is one of the issues to be examined. Of these two limiting operations, we consider restricting attention to small mutation rates to be especially unrealistic. Allowing the mutation rate to be bounded away from zero complicates the analysis but affects neither the techniques nor the basic nature of the results.[19] We consider the assumption of a large population to be the more realistic of the two limits in many applications, though it clearly does not apply to all cases of interest. Once again, we assume throughout this section that Assumptions 1–3 hold.

Two Strict Nash Equilibria. We first assume $A > B$ and $D > C$, so that the game $\mathcal{G}$ has two strict Nash equilibria. As in the previous section, we

[17] Hence, convergence in our model is not fast in the second sense which Ellison [10] discusses because our waiting times do not remain bounded as N gets large.

[18] N often need not be very large before most of the mass of the stationary distribution is attached to a single state. In the example in [7], in which $N = 100$, $z = 33$, and $\lambda = 0.001$, the stationary distribution places more than 0.97 probability on states in which at most 0.05 of the population plays strategy X.

[19] Consider, for example, the case of two strict Nash equilibria. If the mutation rate is positive, then taking the limit as $N \to \infty$ produces a stationary distribution that concentrates all of its probability near one of the strict Nash equilibria, being closer as the mutation rate is smaller. The criterion for determining which equilibrium is "selected" in this way is a variant of (12), with the limits on the integral being adjusted to account for the positive mutation rate.

let $\gamma_{(\lambda, N)}$ denote the limiting stationary distribution of the Markov process on $\{0, 1, ..., N\}$ as $\tau \to 0$. We also use $\gamma_{(\lambda, N)}$ to denote the corresponding Borel measure on $[0, 1]$. Thus, for an open interval $A \subset [0, 1]$, $\gamma_{(\lambda, N)}(A)$ is the probability of finding the system at a state x with $x/N \in A$. To avoid a tedious special case, we assume

$$\int_0^1 (\ln r_0(k) - \ln \ell_0(k))\, dk \neq 0, \tag{9}$$

where Assumptions 1.4 and 2 ensure that the integral exists.

PROPOSITION 4. *Let* (9) *hold. Then there exists a unique Borel probability measure* γ^* *on* $[0, 1]$ *with* $\lim_{N \to \infty} \lim_{\lambda \to 0} \gamma_{(\lambda, N)} = \lim_{\lambda \to 0} \lim_{N \to \infty} \gamma_{(\lambda, N)} = \gamma^*$, *where the limits refer to the weak convergence of probability measures. In addition,* $\gamma^*(0) + \gamma^*(1) = 1$.

Proof. We first calculate $\lim_{N \to \infty} \lim_{\lambda \to 0} \gamma_{(\lambda, N)}$. This becomes our candidate for γ^*.

Fix N. From Assumptions 1.1–1.3, we have

$$\lim_{\lambda \to 0} \frac{r_{(\lambda, N)}(0)}{\ell_{(\lambda, N)}(1)} = \lim_{\lambda \to 0} \frac{\ell_{(\lambda, N)}(N)}{r_{(\lambda, N)}(N-1)} = 0.$$

Using (5) and the fact that $\lim_{\lambda \to 0}(r_{(\lambda, N)}(x)/\ell_{(\lambda, N)}(x+1))$ is nonzero and finite for every value $x \in \{1, 2, ..., N-1\}$ (by Assumptions 1.2–1.3), this result ensures that $\lim_{\lambda \to 0}[\gamma_{(\lambda, N)}(0) + \gamma_{(\lambda, N)}(1)] = 1$. As the mutation rate approaches zero, the system thus spends an increasing amount of time "stuck" at its endpoints, so that in the limit all probability must accumulate on these endpoints.

Hence, we set $\gamma^*(0) + \gamma^*(1) = 1$, and the only remaining question concerns the ratio of these two values. To fix this ratio, we note that for fixed N and λ, we have

$$\frac{\gamma_{(\lambda, N)}(N)}{\gamma_{(\lambda, N)}(0)} = \prod_{x=0}^{N-1} \frac{r_{(\lambda, N)}(x)}{\ell_{(\lambda, N)}(x+1)}.$$

We then take logarithms to obtain

$$\ln \frac{\gamma_{(\lambda, N)}(N)}{\gamma_{(\lambda, N)}(0)} = \sum_{x=0}^{N-1} \{\ln r_{(\lambda, N)}(x) - \ln \ell_{(\lambda, N)}(x+1)\}. \tag{10}$$

The next step is to take the limit of the expression in (10) as $\lambda \to 0$. Because $r_{(0, N)}(0) = \ell_{(0, N)}(N) = 0$, simply replacing λ with 0 on the right side of (10) yields a sum that is undefined, containing one term that equals positive infinity and one that equals negative infinity. However, Assumption 1.5

ensures that the sum of these two terms has a finite limit as $\lambda \to 0$. The remaining terms in (10) can be grouped into pairs, with the two terms of each pair involving the same value of x and with Assumption 1.4 then ensuring that the sum of each pair approaches a finite limit. As a result, we can write

$$\lim_{\lambda \to 0} \ln \frac{\gamma_{(\lambda, N)}(N)}{\gamma_{(\lambda, N)}(0)} = \sum_{x=0}^{N-1} \{\ln r_{(0, N)}(x) - \ln \ell_{(0, N)}(x+1)\},$$

where the right side is defined to be the appropriate limit. Assumptions 2, 1.4 and 1.5 and Lebesgue's dominated convergence theorem ([5, Theorem 16.4]) now give

$$\lim_{N \to \infty} \lim_{\lambda \to 0} \frac{1}{N} \ln \frac{\gamma_{(\lambda, N)}(N)}{\gamma_{(\lambda, N)}(0)} = \int_0^1 (\ln r_0(k) - \ln \ell_0(k))\, dk. \tag{11}$$

Letting our candidate for γ^* satisfy $\gamma^*(0) = 1$, if the right side of (11) is negative, and $\gamma^*(1) = 1$, if the right side of (11) is positive,[20] we then have that $\lim_{N \to \infty} \lim_{\lambda \to 0} \gamma_{(\lambda, N)} = \gamma^*$.[21]

It remains to verify that $\lim_{\lambda \to 0} \lim_{N \to \infty} \gamma_{(\lambda, N)} = \gamma^*$. First, we show that $\lim_{\lambda \to 0} \lim_{N \to \infty} \gamma_{(\lambda, N)}(\{0, 1\}) = 1$. Consider the sets $[k_1, k_2 + \varepsilon]$ and $[k_2, k_2 + \varepsilon]$, for $0 < k_1 < k_2 + \varepsilon < k_2 + k_2 + \varepsilon < k^*$, where k^* is the probability attached to strategy X by the mixed strategy equilibrium. Then for sufficiently small λ, we have

$$\frac{\gamma_{(\lambda, N)}([k_1, k_1 + \varepsilon])}{\gamma_{(\lambda, N)}([k_2, k_2 + \varepsilon])} \geqslant \prod_{x=(k_1+\varepsilon)N}^{k_2 N - 1} \frac{r_{(\lambda, N)}(x)}{\ell_{(\lambda, N)}(x+1)},$$

where (from Assumptions 3.2, 3.3) every term in the product on the right side of this inequality is less than one. Hence, for sufficiently small λ, we have that $\lim_{N \to \infty} \gamma_{(\lambda, N)}([k_2, k_2 + \varepsilon]) = 0$. A similar argument applies to closed subintervals of $[k^*, 1]$ and yields the result.

Next, let (9) be negative. We show that $\lim_{\lambda \to 0} \lim_{N \to \infty} \gamma_{(\lambda, N)}(1 - k, 1] = 0$ for sufficiently small k. Because (9) is negative we can fix k' sufficiently small that $\max_{k'' \leqslant k'} \int_{k'}^{1-k''} (\ln r_0(k) - \ln \ell_0(k))\, dk < 0$. Let $\{N_i\}_{i=1}^{\infty}$ be an increasing sequence of values of N with the property that kN_i is an integer for each i.[22] We then have

$$\frac{\gamma_{(\lambda, N_i)}((1 - k', 1])}{\gamma_{(\lambda, N_i)}(k')} \leqslant k' N_i \max_{x' \in [N_i - k' N_i, N_i]} \prod_{x=k'N_i}^{x'-1} \frac{r_{(\lambda, N_i)}(x)}{\ell_{(\lambda, N_i)}(x+1)}.$$

[20] If the right side of (11) equals zero, then both $\gamma^*(0)$ and $\gamma^*(1)$ may be positive. The limiting arguments are much more tedious in this case, prompting us to invoke (9).

[21] This is a weak convergence claim. By Theorem 2.2 of Billingsley [4], it suffices for weak convergence to show $\lim_{N \to \infty} \lim_{\lambda \to 0} \gamma_{(\lambda, N)}(A) = \gamma^*(A)$ for any relatively open subinterval A of $[0, 1]$, which immediately follows from $\gamma^*(0) + \gamma^*(1) = 1$ and (11).

[22] Generalizing to arbitrary sequences requires somewhat more tedious notation.

This in turn gives

$$\lim_{i\to\infty} \ln \frac{\gamma_{(\lambda, N_i)}((1-k', 1])}{\gamma_{(\lambda, N_i)}(k')}$$

$$\leqslant \lim_{i\to\infty} \left(\ln(k'N_i) + N_i \max_{k''\in[1-k', 1]} \int_{k'}^{1-k''} (\ln r_\lambda(k) - \ln \ell_\lambda(k))\, dk \right),$$

where the integral on the right side is negative. Denoting the value of this integral by c, it suffices to show that $\lim_{N\to\infty} k'Ne^{cN} = 0$. But this follows from l'Hôpital's rule.

An analogous argument for the case in which (9) is positive establishes $\lim_{\lambda\to 0} \lim_{N\to\infty} \gamma_{(\lambda, N)}([0, k)) = 0$ for small k. Hence, for sufficiently small k, $\lim_{\lambda\to 0} \lim_{N\to\infty} \gamma_{(\lambda, N)}$ assigns probability to $[0, k)$ (or $(1-k, 1]$) if and only if γ^* assigns probability to $[0, k)$ (or $(1-k, 1]$), ensuring that $\lim_{\lambda\to 0} \lim_{N\to\infty} \gamma_{(\lambda, N)} = \gamma^*$ in the sense of weak convergence. ∎

We thus have that, in the limit as mutation probabilities get small and the population gets large (in any order), the stationary distribution of the Markov process attaches probability only to the two pure strategy equilibria. In "generic" cases (those for which (9) holds), probability will be attached to only one of these equilibria, which we refer to as the *selected* equilibrium.

Which Equilibrium? A number of papers have recently addressed the problem of equilibrium selection in symmetric 2×2 games. Young [33] and Kandori *et al.* [19] are typical in finding that the risk-dominant equilibrium is always selected. Robson and Vega Redondo [27] offer a model in which the payoff-dominant equilibrium is always selected. However, (11) provides a criterion which shows that our model sometimes selects the payoff-dominant equilibrium and sometimes selects the risk-dominant equilibrium.

COROLLARY 1. [1.1] *The selected equilibrium will be* (X, X) [(Y, Y)] *if*

$$\int_0^1 \ln(r_0(k) - \ln \ell_0(k))\, dk > [<]\ 0. \tag{12}$$

[1.2] *The payoff-dominant equilibrium in game* $\mathcal{G}$ *can be selected even if it fails to be risk dominant.*

Proof. Corollary 1.1 follows immediately from (11). To establish Corollary 1.2, consider the aspiration and imitation model. Let

$$\Delta = 0, \qquad A = 2, \qquad B = 1, \qquad D = 0, \qquad C = -1.$$

Then neither of the two pure strategy Nash equilibria, given by (X, X) and (Y, Y), risk-dominates the other, but (X, X) is the payoff-dominant equilibrium. Let F be a uniform distribution on the interval $[-2, 2]$. Then death probabilities are linear in expected payoffs, with $g(2)=0$, $g(1)=\frac{1}{4}$, $g(0)=\frac{1}{2}$, $g(-1)=\frac{3}{4}$, $g(\pi_Y(k))=\frac{1}{4}k+\frac{1}{2}(1-k)$ and $g(\pi_X(k))=\frac{3}{4}(1-k)$. Inserting these probabilities in (2)–(3), taking the limits $\lambda \to 0$ and $N \to \infty$ and inserting in (11) gives

$$\ln \frac{\gamma^*(1)}{\gamma^*(0)} = \lim_{N \to \infty} N \int_0^1 (\ln g(\pi_Y(k)) - \ln g(\pi_X(k)))\, dk$$

$$= \lim_{N \to \infty} N \int_0^1 (\ln(\tfrac{1}{4}k + \tfrac{1}{2}(1-k)) - \ln(\tfrac{3}{4}(1-k)))\, dk,$$

$$\frac{\gamma^*(1)}{\gamma^*(0)} = \lim_{N \to \infty} \left(\frac{4}{3}\right)^N, \tag{13}$$

ensuring that (X, X) is selected. The game can be perturbed slightly to make (Y, Y) risk-dominant while still keeping (X, X) payoff-dominant without altering the fact that (X, X) is selected. ∎

We can provide some intuition as to why this result differs from that of Kandori *et al.* [19], whose model selects the equilibrium with the larger basin of attraction under best-reply dynamics, namely the risk-dominant equilibrium. In the perturbed version of the game that we considered in the previous proof, the equilibrium (X, X) has a basin of attraction smaller than (Y, Y)'s, but in (X, X)'s basin the death probability of X relative to Y is very small, being nearly zero for states in which nearly all agents play X. This makes it very difficult to leave (X, X)'s basin, and yields a selection in which all agents play X. Only the size of the basin of attraction matters in Kandori *et al.*, while in our model the strength of the learning flows matters as well.

Best-Response Dynamics. The previous paragraph suggests that our muddling model should be more likely to select the risk-dominant equilibrium the closer is the learning process to best-reply learning. We can confirm this.

Let $A > B$ and $D > C$, and let $x^*/N > 1/2$, so that there are two strict Nash equilibria, with (Y, Y) being the risk-dominant equilibrium. Fix $r_0(k)$ and $\ell_0(k)$ satisfying Assumptions 1–3. Then let

$$\tilde{r}_0(k) = \phi B_X(k) + (1-\phi)\, r_0(k)$$

$$\tilde{\ell}_0(k) = \phi B_Y(k) + (1-\phi)\, \ell_0(k),$$

where $B_X(k)$ equals 1 if X is a best response ($k > k^*$) and zero otherwise, and $B_Y(k)$ equals one if Y is a best response ($k < k^*$) and zero otherwise. We say that $\tilde{r}_0(k)$ and $\tilde{\ell}_0(k)$ are a convex combination of the best-response dynamics and $r_0(k)$ and $\ell_0(k)$. As ϕ increases to unity, $\tilde{r}_0$ and $\tilde{\ell}_0$ approach best-response dynamics.

PROPOSITION 5. *If $r_0(k)$ and $\ell_0(k)$ satisfy Assumptions* 1–3, *then the selected equilibrium in game $\mathcal{G}$ is the risk-dominant equilibrium for any convex combination of the best-response dynamics and $r_0(k)$ and $\ell_0(k)$ that puts sufficient weight on the former.*

Proof. Let $k^* \equiv x^*/N > 1/2$, so that (Y, Y) is the risk-dominant equilibrium. Condition (12) will fail for sufficiently large ϕ, and hence the selected equilibrium will be (Y, Y), if

$$\lim_{\phi \to 1} \int_0^{k^*} \{\ln((1-\phi)\, r_0(k)) - \ln(\phi + (1-\phi)\, \ell_0(k))\}\, dk$$

$$+ \int_{k^*}^1 \{\ln(\phi + (1-\phi)\, r_0(k)) - \ln((1-\phi)\, \ell_0(k))\}\, dk$$

$$= \lim_{\phi \to 1} \int_0^{k^*} \{\ln(r_0(k)) - \ln(\phi + (1-\phi)\, \ell_0(k))\}\, dk \tag{14}$$

$$+ \int_{k^*}^1 \{\ln(\phi + (1-\phi)\, r_0(k)) - \ln(\ell_0(k))\}\, dk \tag{15}$$

$$+ \ln(1-\phi)\left(\int_0^{k^*} dk - \int_{k^*}^1 dk\right) < 0. \tag{16}$$

The sum of terms (14) and (15) approaches a finite number as ϕ approaches unity. Because (Y, Y) is risk dominant, $k^* > \frac{1}{2}$ and (16) approaches negative infinity, and hence the result. ∎

No Pure Strategy Equilibria. Our equilibrium selection results address the case of two strict Nash equilibria. We can contrast these results with the case of games in which $B > A$ and $C > D$, so that there is a unique, mixed-strategy Nash equilibrium. Then an argument analogous to that of Proposition 4 yields:

PROPOSITION 6. *Let k^* be the probability attached to X in the mixed-strategy equilibrium. Then $\lim_{\lambda \to 0} \lim_{N \to \infty} \gamma_{(\lambda, N)}(A) = 0$ if $k^* \notin A$. However, $\lim_{\lambda \to 0} \gamma_{(\lambda, N)}(0) + \lim_{\lambda \to 0} \gamma_{(\lambda, N)}(1) = 1$.*

The order of limits makes a difference in this case. If mutation rates are first allowed to approach zero, then the ultralong-run dynamics are driven by the possibility of accidental extinction coupled with the impossibility of recovery, attaching probability only to the two nonequilibrium states in which either all agents play X or all agents play Y. If the population size is first allowed to get large, then accidental extinctions are not a factor and the selected outcome is the mixed strategy equilibrium. Our inclination here is to regard the latter as the more useful model.

5. RISK-DOMINANCE

The comparative statics results of the previous section involve changes in the learning rule. We now fix a learning rule and examine changes in the payoffs of the game.

Some additional assumptions are required to establish comparative static results. Assumptions 1–3 are silent on the question of how changes in payoffs affect the learning dynamics as long as the inequalities in Assumption 3 are satisfied. We accordingly consider the aspiration and imitation model. We investigate games with two strict Nash equilibria ($A > B$ and $D > C$) and ask when the risk-dominant equilibrium will be selected. The support of the random variable $\Delta - \tilde{R}$ is assumed to be a closed interval encompassing A, B, C, and D.

We begin with the case in which there is no conflict between payoff and risk-dominance:

PROPOSITION 7. *If the payoff-dominant equilibrium in game $\mathcal{G}$ is also risk dominant, then the payoff-dominant equilibrium is selected.*

Proof. In the aspiration and imitation model, the criterion given by (12) for the selection of equilibrium (X, X) becomes

$$\int_0^1 (\ln F(\Delta - \pi_Y(k)) - \ln F(\Delta - \pi_X(k)))\, dk$$

$$= \int_0^1 \ln F(\Delta - kB - (1-k)\, D)\, dk - \int_0^1 \ln F(\Delta - kA - (1-k)\, C)\, dk > 0. \tag{17}$$

Let (X, X) and (Y, Y) be risk-equivalent in game $\mathcal{G}$, so that $A + C = B + D$, and let $A = D$. Then (17) holds with equality. Now make (X, X) the payoff-dominant equilibrium by increasing A and decreasing C by a like amount, so as to preserve $A + C = B + D$ (and hence to preserve the risk-equivalence

of the two strategies). Because ln F is concave, this mean-preserving spread in the interval of values over which ln F is integrated decreases the second integral in (17). The expression (17) then becomes positive and the payoff-dominant equilibrium (X, X) is selected. Next, note that adding a constant to A and C or subtracting a constant from D and B so as to also make (X, X) risk-dominant increases (17) and hence preserves the result that the payoff-dominant equilibrium is selected. ▮

We now consider cases where the payoff- and risk-dominance criteria conflict. Let (Y, Y) be the risk-dominant equilibrium, so k^*, the probability attached to X by the mixed-strategy equilibrium, exceeds $\frac{1}{2}$, but let (X, X) be payoff-dominant. Let π^* be the payoff in game $\mathscr{G}$ from the mixed-strategy equilibrium. We will consider variations in the payoffs A, B, C, D that leave k^* and π^* unchanged. For example, we will consider a decrease in D accompanied by an increase in B calculated so as to preserve k^* and π^*, as illustrated in Fig. 3. We thus restrict attention to variations in the payoffs A, B, C, and D for which $C = C(A)$ and $B = B(D)$, where $(1-k^*)\,C(A) + k^*A = k^*B(D) + (1-k^*)\,D = \pi^*$.

PROPOSITION 8. *If the payoff-dominant equilibrium is selected given payoffs A, B, C and D, with mixed-strategy equilibrium k^* and mixed-strategy equilibrium payoff π^*, and if $D > B$, then there exists $\underline{D} < \pi^*$ such that the payoff-dominant equilibrium is also selected for any payoffs A, B', C, and D' that preserve k^* and π^* and satisfy $D' \in [D, \underline{D})$.*

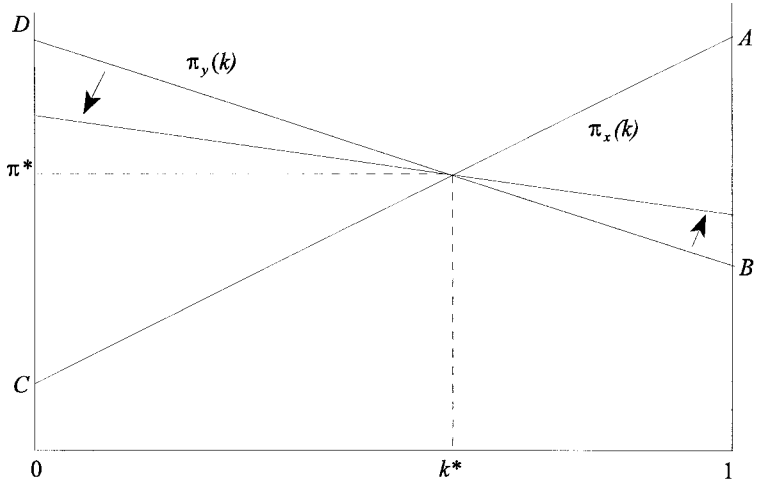

FIG. 3. Payoff variations.

Proposition 8 tells us that if the payoff-dominant equilibrium is selected for payoffs A, B, C, and D with $D > B$, then the payoff-dominant equilibrium is selected for an interval of values of D' (with B' satisfying $k^*B' + (1-k^*)\,D' = \pi^*$) containing D and containing π^* in its interior. However, if we fix A, C, π^*, and k^* and find that the payoff-dominant equilibrium is selected for no values of D and B with $D > \pi^*$, it may still be that the payoff-dominant equilibrium is selected for some values of $D < \pi^*$ (and hence $D < B$). The best case for the payoff-dominant equilibrium thus occurs when either $D = B$ or $D < B$. The payoff-dominant equilibrium is favored by reducing the variation in payoffs to strategy Y or even "inverting" them, so that while (Y, Y) is an equilibrium, the highest reward to strategy Y is obtained if the opponent plays X.

Proof of Proposition 8. Fix $k^* > \frac{1}{2}$. Fix A and hence $C(A)$. If we set $D = A$, then Proposition 7 ensures that (Y, Y) will be selected, since it is risk dominant and payoff undominated. Now let D decline and B increase at slower rate (to preserve $k^*B + (1-k^*)\,D = \pi^*$ with $k^* > \frac{1}{2}$). Taking the derivative of (17), we find that because $\ln F$ is concave, (17) increases until D reaches π^* and is increasing at $D = \pi^* = B$. Hence, if the payoff-dominant equilibrium is selected for any value of $D > \pi^*$, then it is also selected for any smaller value $D \geqslant \pi^*$ and for some values of $D < \pi^*$. ∎

Some Experimental Evidence. Given Proposition 8, it is interesting to note that Straub [31] has conducted experiments to investigate the conditions under which risk-dominant and payoff-dominant equilibria are selected in 2×2 symmetric games with two strict Nash equilibria. He finds that the risk-dominant equilibrium is the most common equilibrium played in seven out of eight of the experiments. The exception, in which the payoff-dominant equilibrium appeared, is the only game in which $D < B$. Friedman [13] also reports experiments with 2×2 symmetric games with two strict Nash equilibria. Friedman finds that the payoff-dominant equilibrium is chosen more often in a game with $D < B$ than in a game with the same values of A and C and the same basins of attraction (same value of k^* and π^*) but with D and B altered to $D = B$. Both experiments are consistent with Proposition 8.

6. ENDOGENOUS LEARNING

The heart of our model is a learning rule. Which rules are worthy of our attention? A useful way to approach this question is to recognize that learning rules themselves are likely to have been acquired through an evolutionary process.

We capture the evolution of learning rules in a "two-tiered" model. We view the evolution of strategy choices, guided by a particular learning rule, as proceeding at a pace that is rapid compared to the evolution of the learning rule itself. We take our existing model to represent the evolution of strategy choices given a learning rule. The payoffs received from the strategy choices that emerge from this process then drive the evolution of learning rules.

An attempt to model the evolution of learning rules raises the specter of an infinite regress. A model in which agents learn how to play games now becomes a model in which agents learn how to learn to play games. But then why not a model in which agents learn how to learn how to learn, and so on? We might escape the infinite regress if outcomes are not particularly sensitive to the nature of the learning rules that agents use in learning how to learn, so that there is no need for any learning past the second stage. Toward this end, we look for cases in which learning rules satisfy conditions analogous to evolutionary stability.[23] If this robustness had appeared at the first level, when examining how agents learn to play games, we would not have been prompted to seek a level higher.

We again restrict attention to the aspiration and imitation model, with the random variable $\tilde{R}$ assumed to have a strictly positive density on the real line. In addition, we allow only the value of the aspiration level Δ to be subject to change.[24] We therefore label a learning rule by the aspiration level Δ which it incorporates and we refer to the evolution of Δ as the evolution of a learning rule. Only games with two strict Nash equilibria are considered. The proof of the following is contained in the Appendix:

PROPOSITION 9. *Let $\Delta' < \Delta$. Then for sufficiently large N and sufficiently small λ, the payoff to a player characterized by aspiration level Δ', in any population consisting of aspiration levels Δ and Δ', exceeds the payoff to a player characterized by Δ.*

The mechanism behind this result is straightforward. In any stationary distribution, players spend long periods of time facing a mix of strategies that is concentrated on a particular strategy (say Y) but also includes other strategies. The highest expected payoffs will be garnered by those agents whose learning rules cause them to spend the greatest proportion of the time playing the best reply Y. These will be agents with learning rules that

[23] See Harley [18], Maynard Smith [22], and Ellison and Fudenberg [11] for work in this vein.

[24] We view the information available to agents and the distribution of $\tilde{R}$ as being part of the technology of the game. It would be interesting to consider models in which players might take steps, perhaps at a cost, to influence this latter distribution. Bendor *et al.* [3] suggest that aspiration levels should adjust to equal the average equilibrium payoff. In Binmore and Samuelson [6], we show that this is not always possible in a muddling model.

make them relatively unlikely to switch away from high payoff realizations and relatively likely to switch away from low payoff realizations. In the aspiration and imitation model with log-concave F, these learning rules involve smaller aspiration levels.

We thus have a tendency for evolution to select smaller aspiration levels. What are the implications of smaller aspiration levels for the selected equilibrium? Here we specialize to the standard normal distribution.[25]

PROPOSITION 10. *Let F be the standard normal distribution. Then for sufficiently small Δ, the selected equilibrium is risk-dominant.*

Proof. Let $A > C$ and $D > B$ with $A + C < B + D$, so that (Y, Y) is risk-dominant. It follows from (12) that a sufficient condition for the risk-dominant equilibrium to be selected is that $\ln F(\Delta - B) + \ln F(\Delta - D) - \ln F(\Delta - A) - \ln F(\Delta - C) < 0$ hold for *any* such A, B, C and D. Let $N_A = 1 - F(-(\Delta - A))$, and let N_B, N_C, and N_D be defined analogously. Then it suffices to show that, for small values of Δ,

$$\frac{N_B N_D}{N_A N_C} < 1.$$

For the standard normal distribution, l'Hôpital's rule can be used to show that $\lim_{s \to \infty} N_{s+z}/N_s = e^{-z^2/2} e^{-sz}$. Hence, we need to show that, for small values of Δ (and hence large values of $-\Delta$),

$$\frac{e^{-(D-C)^2/2} e^{-(D-C)(-\Delta + C)}}{e^{-(A-B)^2/2} e^{-(A-B)(-\Delta + B)}} < 0.$$

This in turn is equivalent to showing that $(A - B)[(A - B) + 2(-\Delta + B)] - (D - C)[(D - C) + 2(-\Delta + C)] < 0$. As $-\Delta$ gets large, we need only examine the terms involving $-\Delta$, which gives $2(-\Delta)(A + C - B - D) < 0$. This will hold for large $-\Delta$ because (Y, Y) is risk-dominant, so that $A + C - B - D < 0$ and hence the coefficient of $-\Delta$ is negative. ∎

We therefore have identified cases in which evolution will tend to select learning rules that lead to the risk-dominant equilibrium. However, we have examined the evolution in the context of a very narrow class of learning rules, namely the aspiration and imitation model where F is the normal distribution. Within this class of rules we have allowed evolution to affect only the aspiration level. Furthermore, we have identified cases in which lower aspiration levels fare better than higher aspiration levels but have not

[25] The normal distribution satisfies our assumption that $\ln F$ is concave.

explicitly modeled the process by which different aspiration levels appear and contend for survival. What happens in more general cases remains open.

7. CONCLUSION

Evolutionary game theory offers the promise of progress on the problem of equilibrium selection. At the same time, it is capable of reproducing the worst features of the equilibrium refinements' literature, creating an ever-growing menagerie of conflicting and uninterpreted results. To achieve the former rather than the latter outcome, we think that evolutionary models need to be provided with microfoundations which identify the links between the dynamics of the model and the underlying choice behavior.

In this paper, we focus on an aspect of choice behavior that we consider particularly important: we allow people to make mistakes in choosing their strategies. Ours is thus a muddling rather than a maximizing model, with the primary source of noise in our model being nonnegligible mistakes within the learning process itself. Introducing muddling behavior has implications both for equilibrium selection (where we find that the payoff-dominant equilibrium is sometimes selected) and also for questions of timing. In particular, we find that the length of time needed to reach the ultralong run may be shorter in a muddling than in a maximizing model, making it more likely that the ultralong run will be of interest in potential applications.

The paper closes with a model in which the rules by which agents learn to play games are themselves subject to evolutionary pressures. Our work here is both preliminary and incomplete. But we believe this to be an important area for further work.

APPENDIX: PROOFS

Proof of Proposition 3. Since we will be working in the limit as the length of a time period τ becomes arbitrarily short, we can assume that in each time period of length τ, either no agent receives the learn-draw (with probability $1-\tau N$) or a single agent receives the learn-draw (with probability τN). Let $\hat{\Gamma}_{(\lambda, N, \tau)}$ be the resulting Markov process, let $\hat{\gamma}_{(\lambda, N, \tau)}$ be its stationary distribution and notice that $\hat{\gamma}_{(\lambda, N, \tau)} = \gamma_{(\lambda, N)}$.

Fix a time t and a period length τ, so that t/τ periods will have occurred by time t. Let $\iota(k, z)$ be the probability that out of z periods, there are exactly k periods in which some individual receives the learn-draw. Then we have

$$\gamma^0[\hat{\Gamma}_{(\lambda, N, \tau)}]^{t/\tau} = \sum_{k=0}^{t/\tau} \iota\left(k, \frac{t}{\tau}\right)[\Gamma^*_{(\lambda, N)}]^k,$$

where $\Gamma^*_{(\lambda, N)}$ is the transition matrix contingent upon one learn-draw having been received and we take $[\Gamma^*_{(\lambda, N)}]^0$ to be the identity matrix. Hence, it suffices for (6) to show that for any γ^0,

$$\lim_{\tau \to 0} \left(1 - \left\| \gamma^0 \sum_{k=0}^{t/\tau} \iota\left(k, \frac{t}{\tau}\right) [\Gamma^*_{(\lambda, N)}]^k - \hat{\gamma}_{(\lambda, N, \tau)} \right\|^{1/(t-1)}\right) \leqslant h_\gamma(\lambda) \sim \lambda. \tag{18}$$

Let $\tau_n = 1/n$ for $n \in \{1, 2, \ldots\}$. Then $t/\tau_n = nt$, an equality we shall use repeatedly. For each n, let $\{Z_{nh}, h \in \{1, ..., nt\}\}$ be a collection of random variables, one for each of the periods that occur by time t, with each random variable taking the value one if a learn-draw is received (with probability $\tau_n N$) and zero if a learn-draw is not received (with probability $1 - \tau_n N$). Then $\iota(k, nt)$ is the probability that exactly k of the random variables $\{Z_{nh}, h \in \{1, ..., nt\}\}$ take the value one. Notice that, for any n, we have $\sum_{k=1}^{nt} \tau_n N = nt\tau_n N = tN$, to that for any n, the sum over the collection $\{Z_{nh}, h \in \{1, ..., nt\}\}$ of the probabilities of receiving the outcome one is finite and given by tN. Coupled with the fact that τ_n and $\tau_n N$ approach zero as n gets large, this allows us to apply Theorem 23.2 of Billingsley [5] to conclude that

$$\lim_{n \to \infty} \iota\left(k, \frac{t}{\tau_n}\right) = \frac{(Nt)^k}{k!} e^{-Nt}.$$

Hence, as τ_n gets small, $\iota(k, t/\tau_n) = \iota(k, nt)$ is given by a Poisson distribution with mean and variance Nt. It accordingly suffices for (18) to show, for any γ^0, that

$$\left\| \gamma^0 \sum_{k=0}^{\infty} \frac{(Nt)^k}{e^{Nt} k!} [\Gamma^*_{(\lambda, N)}]^k - \gamma_{(\lambda, N)} \right\| \leqslant (1 - h_\gamma(\lambda))^{t-1}. \tag{19}$$

We first observe that $\hat{\gamma}_{(\lambda, N, \tau)} \hat{\Gamma}_{(\lambda, N, \tau)} = \hat{\gamma}_{(\lambda, N, \tau)}((1 - \tau N) I + \tau N \Gamma^*_{(\lambda, N)}) = \hat{\gamma}_{(\lambda, N, \tau)}$, which we can solve and then take the limit $\tau \to \infty$ to obtain $\gamma_{(\lambda, N)} \Gamma^*_{(\lambda, N)} = \gamma_{(\lambda, N)} I = \gamma_{(\lambda, N)}$. Then $\gamma_{(\lambda, N)}$ is the (unique) stationary distribution of the matrix $[\Gamma^*_{(\lambda, N)}]$, and so

$$\lim_{k \to \infty} \gamma^0 [\Gamma^*_{(\lambda, N)}]^k = \gamma_{(\lambda, N)}.$$

The matrix $[\Gamma^*_{(\lambda, N)}]$ has many zero elements, but the matrix $[\Gamma^*_{(\lambda, N)}]^N$ is strictly positive. Corollary 4.1.5 of Kemeny and Snell ([21], page 71) can therefore be applied to show that

$$\| \gamma^0 ([\Gamma^*_{(\lambda, N)}]^N)^t - \gamma_{(\lambda, N)} \| \leqslant (1 - 2S(\lambda))^{t-1}, \tag{20}$$

Where $S(\lambda)$ is the smallest transition probability in $[\Gamma^*_{(\lambda, N)}]^N$. We must then examine the probability $S(\lambda)$. It is not immediately obvious which is the least likely transition in the matrix $[\Gamma^*_{(\lambda, N)}]^N$. One possibility is that it is the transition from the state in which all agents play Y ($x=0$) to the state in which all agents play X ($x=N$). If so, then $S(\lambda)$ is given by

$$\prod_{x=0}^{N-1} r_{(\lambda, N)}(x, N) = \lambda[c' + \theta(\lambda)], \tag{21}$$

where c' does not depend on λ and $\lim_{\lambda \to 0} \theta(\lambda) = 0$, and where the equality follows because for all $x \in \{1, 2, ..., N-1\}$, $\lim_{\lambda \to 0} r_{(\lambda, N)} = r_{(0, N)} > 0$. For sufficiently small λ, we have $|\theta(\lambda)| < \varepsilon$ for some $\varepsilon > 0$. We then let $c = c' - \varepsilon$ and $C = c' + \varepsilon$ to obtain $S(\lambda) \sim \lambda$. A similar argument establishes that *any* transition within $[\Gamma^*_{(\lambda, N)}]^N$ can be made with at most one step that requires a mutation, ensuring that $S(\lambda) \sim \lambda$.

An argument analogous to that leading to (20) gives, for any $\alpha > 0$,

$$\|\gamma^0[\Gamma^*_{(\lambda, N)}]^{\alpha Nt} - \gamma_{(\lambda, N)}\| \leqslant (1 - 2S(\lambda))^{\alpha t - 1}.$$

This would allow us to conclude that there exists a function $h_\gamma(\lambda) \sim \lambda$ such that

$$\limsup_{t \to \infty} \left\| \gamma^0 \sum_{k=0}^{\infty} \frac{(Nt)^k}{e^{Nt} k!} [\Gamma^*_{(\lambda, N)}]^k - \gamma_{(\lambda, N)} \right\|^{1/(\alpha t - 1)} \leqslant (1 - h_\gamma(\lambda)) \tag{22}$$

holds for $\alpha \in (0, 1)$ if we could conclude that

$$\lim_{t \to \infty} \gamma^0 \frac{(Nt)^k}{e^{Nt} k!} [\Gamma^*_{(\lambda, N)}]^k = 0 \tag{23}$$

for any $k < \alpha Nt$. This in turn follows from the well-known observation that as t grows, the Poisson distribution with mean and variance Nt approaches a normal distribution with mean and variance Nt (e.g., Billingsley [5, Problem 27.3 on p. 379]). Equation (23) is then $\lim_{t \to \infty} \text{prob}[\text{N}(Nt, (Nt)^{1/2}) < \alpha Nt] = \lim_{t \to \infty} \text{prob}[\text{N}(0, 1) < (\alpha - 1)(Nt)^{1/2}] = 0$, which follows from the fact that $(\alpha - 1)(Nt)^{1/2} \to -\infty$ as $t \to \infty$.

Finally, we note that because (22) holds for any $\alpha \in (0, 1)$, we must have (19), which is the desired result. ∎

Proof of Proposition 9. Let the agents in the population be distributed between the aspiration levels Δ and Δ' with $\Delta' < \Delta$. For sufficiently large population size and small mutation rates, there exist numbers $p_X(N, \lambda)$, $p_Y(N, \lambda)$, $x_Y(N, \lambda)$ and $x_X(N, \lambda)$, with the sum of the first two numbers arbitrarily close to one and the latter two numbers arbitrarily lose to 0 and

1 respectively, such that the stationary distribution induced by the prevailing collection of learning rules spends a proportion of at least $p_Y(N, \lambda)$ of the time in states in which $x/N < x_Y(N, \lambda)$ (in which case Y is a best reply) and $p_X(N, \lambda)$ of the time in states in which $x/N > x_X(N, \lambda)$ (in which case X is the best reply). Call these sets of states P_Y and P_X, and call the remaining states P_D.

The difference between the payoffs to aspiration levels Δ' and Δ is

$$p_Y(N, \lambda)\, \Pi(P_Y) + p_X(N, \lambda)\, \Pi(P_X) + (1 - p_Y(N, \lambda) - p_X(N, \lambda))\, \Pi(P_D),$$

where $\Pi(P_Y)$ is the expected payoff difference between aspiration levels Δ' and Δ conditional on the system being in the set P_Y, and $\Pi(P_X)$ and $\Pi(P_D)$ are defined analogously. Because the time spent in the set P_D can be made arbitrarily small by increasing N and decreasing λ, it suffices to show that $p_Y(N, \lambda)\, \Pi(P_Y)$ or $P_X(N, \lambda)\, \Pi(P_X)$ are positive, and at least one is bounded away from zero as N increases and λ decreases.

At least one of $p_Y(N, \lambda)$ and $p_X(N, \lambda)$ must be bounded away from zero. Suppose it is $p_Y(N, \lambda)$. Then

$$\Pi(P_Y) = \sum_{k=0}^{\infty} \left(\frac{k p_Y^k(N, \lambda)}{\sum_{h=0}^{\infty} h p_Y^h(N, \lambda)} \right) \Pi^k(P_Y), \tag{24}$$

where $p_Y^k(N, \lambda)$ is the probability that a given episode during which the system is in the set P_Y lasts k periods, and $\Pi^k(P_Y)$ is the expected per-period payoff difference between learning rules Δ' and Δ during a collection of periods in which the system stays in the set P_Y for exactly k periods. Then for sufficiently small λ we have

$$\lim_{N \to \infty} \sum_{h=0}^{\infty} h p_Y^h(N, \lambda) = \infty, \tag{25}$$

because, as N gets large, the system spends an arbitrarily small proportion of its time in P_D and every stay in P_Y must end with an entry into P_D.

Next, let $\Pi^\infty(P_Y)$ be the difference in the payoffs to learning rules with aspiration levels Δ' and Δ, conditional on the system staying in the set P_Y an *infinite* number of periods. As the length of a stay in the set P_Y increases, the difference in payoffs between aspiration levels Δ' and Δ, contingent on such a stay, must approach $\Pi^\infty(P_Y)$. Assume temporarily that $\Pi^\infty(P_Y) > 0$. Then for any $\varepsilon > 0$, there is a $\delta' > 0$ such that for all $\delta > \delta'$,

$$\Pi^\delta(P_Y) > \Pi^\infty(P_Y) - \varepsilon. \tag{26}$$

Then (26) and (25) (along with $\Pi^{\infty}(P_Y) > 0$) imply the desired result that (24) is positive for sufficiently large N and small enough λ, and does not approach zero as N grows and λ shrinks.

It then remains only to show that $\Pi^{\infty}(P_Y)$ is positive when $\Delta' < \Delta$. For this, however, it suffices that $F(\Delta - z)/F(\Delta)$ is increasing in Δ for any $z > 0$. In particular, for every state in the set P_Y, Y is a best reply and X is an inferior reply. In addition, by making N sufficiently large, P_Y can be made to allocate enough of its probability to such a small set that the lowest payoff to a best reply over states in this set exceeds the highest payoff to an inferior reply and this difference is arbitrarily large relative to the expected difference in payoffs to a best reply or to an inferior reply over the set P_Y. We can then think of the payoffs to each agent as being determined by the stationary distribution of a two-state Markov process, with the two states being "best reply" and "inferior reply," and with the latter giving a higher payoff than the former. Call this Markov process Γ^*. If $F(\Delta - z)/F(\Delta)$ is increasing in Δ, then the ratio of the probability of abandoning a best reply for an inferior reply to the probability of abandoning an inferior for a best reply is lower for aspiration level Δ' then for Δ. This in turn implies that the stationary distribution of Γ^* must spend more time in the best-reply state for learning rule Δ' than for Δ. Then the former must then receive a higher payoff, yielding the result. Finally, we then need only note that $F(\Delta - z)/F(\Delta)$ will be increasing in Δ if $f(\pi)/F(\pi)$ is decreasing in π, which is equivalent to the log-concavity of F. ▌

REFERENCES

1. M. Amir and S. Berninghaus, Another approach to mutation and learning in games, *Games Econ. Behav.* **14** (1995), 19–43.
2. M. Bagnoli and T. Bergstrom, Log-concave probability and its application, mimeo, Univ. of Michigan, 1989.
3. J. Bendor, D. Mookherjee, and D. Ray, Aspirations, adaptive learning and cooperation in repeated games, mimeo, Indian Statistical Institute, 1994.
4. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.
5. P. Billingsley, "Probability and Measure," Wiley, New York, 1986.
6. K. Binmore and L. Samuelson, "Muddling Through: Noisy Equilibrium Selection," SSRI working paper 9410, University of Wisconsin, 1993.
7. K. Binmore, L. Samuelson, and R. Vaughan, Musical chairs: Modelling noisy evolution, *Games Econ. Behav.* **11** (1995), 1–35.
8. L. E. Blume, The statistical mechanics of strategic interaction, *Games Econ. Behav.* **5** (1993), 387–424.
9. R. R. Bush and F. Mosteller, "Stochastic Models for Learning," Wiley, New York, 1955.
10. G. Ellison, Learning, local interaction, and coordination, *Econometrica* **61** (1993), 1047–1072.
11. G. Ellison and D. Fudenberg, "Rules of Thumb for Social Learning," IDEI discussion paper 17, University of Toulouse, 1992.

12. M. I. Friedlin and A. D. Wentzell, "Random Perturbations of Dynamical Systems," Springer-Verlag, New York, 1984.
13. D. Friedman, Equilibrium in evolutionary games: Some experimental results, *Econ. J.* **106** (1996), 1–25.
14. D. Fudenberg and C. Harris, Evolutionary dynamics with aggregate shocks, *J. Econ. Theory* **57** (1992), 420–441.
15. I. Gilboa and D. Schmeidler, Case-based consumer theory, mimeo, Northwestern University, 1993.
16. I. Gilboa and D. Schmeidler, Case-based decision theory, *Quart. J. Econ.* **110** (1995), 605–640.
17. I. Gilboa and D. Schmeidler, Case-based optimization, *Games Econ. Behav.* (1996), in press.
18. C. B. Harley, Learning the evolutionarily stable strategy, *J. Theor. Biol.* **89** (1981), 611–633.
19. M. Kandori, G. J. Mailath, and R. Rob, Learning, mutation, and long run equilibria in games, *Econometrica* **61** (1993), 29–56.
20. S. Karlin and H. M. Taylor, "A First Course in Stochastic Processes," Academic Press, New York, 1975.
21. J. G. Kemeny and J. L. Snell, "Finite Markov Chains," Van Nostrand, Princeton, NJ, 1960.
22. J. M. Smith, "Evolution and the Theory of Games," Cambridge Univ. Press, Cambridge, UK, 1982.
23. P. R. Milgrom, Good news and bad news: Representation theorems and applications, *Bell J. Econ.* **12** (1981), 380–391.
24. R. Nelson and S. Winter, "An Evolutionary Theory of Economic Change," Harvard Univ. Press, Cambridge, MA, 1982.
25. G. Nöldeke and L. Samuelson, An evolutionary analysis of backward and forward induction, *Games Econ. Behav.* **5** (1993), 425–454.
26. A. J. Robson, A biological basis for expected and non-expected utility, *J. Econ. Theor.* **68** (1996), 397–424.
27. A. J. Robson and F. Vega-Redondo, Efficient equilibrium selection in evolutionary games with random matching, *J. Econ. Theor.* **70** (1995), 65–92.
28. H. Simon, A behavioral model of rational choice, *Quart. J. Econ.* **69** (1955), 99–118.
29. H. Simon, "Models of Man," Wiley, New York, 1957.
30. H. Simon, Theories of decision making in economics and behavioral science, *Amer. Econ. Rev.* **49** (1959), 253–283.
31. P. G. Straub, "Risk Dominance and Coordination Failures in Static Games," Dispute Resolution Research Center Working Paper 106, Northwestern Univ., 1993.
32. S. Winter, Satisficing, selection and the innovating remnant, *Quart. J. Econ.* **85** (1971), 237–260.
33. P. Young, The evolution of conventions, *Econometrica* **61** (1993), 57–84.