WI

94-10 R

UNIVERSITY
OF
WISCONSIN-
MADISON

SSRI

Muddling Through: Noisy Equilibrium Selection[1]

Ken Binmore
Larry Samuelson

9410R

Workshop Series

SOCIAL SYSTEMS RESEARCH INSTITUTE

MUDDLING THROUGH: NOISY EQUILIBRIUM SELECTION[1]

Ken Binmore
Department of Economics
University College London
Gower Street
London WC1E 6BT England

Larry Samuelson
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, Wisconsin 53706 USA

February 9, 1995

[1]First draft August 26, 1992. Financial support from 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, Reinhard Selten, Avner Shaked, Richard Vaughan, an associate editor and two anonymous referees for helpful discussions and comments.

Abstract

We examine an evolutionary model in which the primary source of "noise" that moves the model between equilibria is not random, arbitrarily improbable mutations but mistakes in learning. We find conditions 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. We also find that waiting times until the limiting distribution is reached can be shorter than in a mutation-driven model. We present comparative static results as well as a "two-tiered" evolutionary model in which the rules by which agents learn to play the game are themselves subject to evolutionary pressure.

Journal of Economic Literature Classification Number C70.

Keywords: Equilibrium selection, Evolutionary games, Mutations, Risk dominance.

MUDDLING THROUGH: NOISY EQUILIBRIUM SELECTION

by Ken Binmore and Larry Samuelson

Commonsense is a method of arriving at workable conclusions from false premisses 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 in which the strategy-adjustment process is explicitly modeled.

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 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 very fact that varying the details in a dynamic model can alter the equilibrium selected shows that the institutional environment in which a game is learned and played can matter for equilibrium selection. Theories of equilibrium selection therefore 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, Samuelson and Vaughan [6] we examined 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 agents 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 terms of extracting ultralong-run equilibrium selection results from explicit learning models are Young [33] and Kandori, Mailath and Rob [19]. In their models, 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 $\lambda \to 0$. The striking prediction of both 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 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 usually work, but arise quite naturally in the noisy world in which games are actually played, where agents may find it difficult even to identify the set of available strategies and the payoffs that these strategies bring.

Examining muddlers rather than maximizers has the consequence that the expected waiting time before the ultralong-run predictions of the model become relevant is 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 Young [33] and Kandori, Mailath and Rob [19], a large number of *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.

[1] Kandori, Mailath and Rob [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. Although we consider long-run phenomena at least as important as the ultralong-run, we concentrate on the latter in this paper.

Incorporating noisy learning into the model also has implications for equilibrium selection: muddling models do not always select the same equilibria as maximizing models. In the symmetric 2×2 games studied in this paper, maximizing models always 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. There are therefore grounds for directing suspicion at risk-dominance as a refinement of Nash equilibrium even in symmetric 2×2 games.

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 imposes some additional structure on the learning process. We then derive conditions under which the payoff-dominant or risk-dominant equilibrium will be selected. These conditions yield testable predictions.

Section 6, with the help of considerably 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 the evolutionarily stable rules from a particular class of learning rules in our muddling model select the risk-dominant equilibrium for symmetric 2×2 games, thus matching the results of maximizing models.

2 A Muddling Model

The Game. We begin with the symmetric 2×2 game $\mathcal{G}$ of Figure 1. We will refer to the entries in this matrix as *expected payoffs*, where these are the familiar von Neumann and Morgenstern utilities of conventional game theory, and where the randomness that motivates the label "expected" will be introduced shortly.

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

	X	Y
X	A, A	C, B
Y	B, C	D, D

Figure 1: The game $\mathcal{G}$

is characterized by the strategy X or Y that she is programmed to use in that period. 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 with replacement to play the game, 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.[2] (We also examine a special case below in which the model is formally identical to one in which each agent plays only once in each period.)

An agent playing X receives an expected payoff of A in a population in which all agents play X and an "average" expected payoff of $\pi_X(k) = kA+(1-k)C$ when a proportion k of her opponents play X and a proportion $(1-k)$ play Y. She receives $\pi_Y(k) = kB + (1-k)D$ when playing Y under similar circumstances. Realized payoffs are random, being given by the average expected payoff in the game $\mathcal{G}$ plus the outcome R of a random variable $\tilde{R}$. We view $\tilde{R}$ as capturing a variety of random shocks that perturb payoffs.[3] We think this randomness is a crucial feature of many real-world

[2]Since we consider the case when $\tau \rightarrow 0$, this assumption has the effect 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 the natural setting for evolutionary models. Kandori, Mailath and Rob [19] assume that agents play an infinite number of times in each period. Nöldeke and Samuelson [25] assume a round-robin tournament in each period. 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].

[3]The random variable $\tilde{R}$ yields a shock common to *each* payoff received by an agent in

games, where players may encounter difficulties even identifying their payoffs precisely. This randomness in turn may be an important reason why learning often proceeds in a muddling rather than maximizing fashion.

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 below. The special case also satisfies Assumptions 4–5.

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 studied by taking the limit $t \longrightarrow \infty$. We then take the limit $\tau \longrightarrow 0$. This gives a model in which agents revise their strategies at uncoordinated, idiosyncratic times.[4] Finally, we take the limits $N \longrightarrow \infty$ and $\lambda \longrightarrow 0$ in order to sharpen the results. We comment on the order of these last two limits as we proceed. An extended discussion of the implications of taking limits in different orders appears in Binmore, Samuelson and Vaughan [6].

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 with probability $\beta\tau$. (Without loss of generality, we take $\beta = 1$.) 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 $\tau \longrightarrow 0$, 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

the given period. The distribution F of $\tilde{R}$ is independent and identically distributed across players choosing the same strategy, but may depend on the current state or strategy. It would also be interesting to study cases 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, or in which the distribution of $\tilde{R}$ differs across players. Papers in which the former type of uncertainty appears include Fudenberg and Harris [14] and Robson [26].

[4] Other models are also worthy of attention. In markets were information regularly becomes publicly available, for example, strategy revisions may be simultaneously undertaken by many or all agents.

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 playing strategy X to switch to Y. (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 Y to switch to X. (Hence $\ell_{(\lambda,N)}(0) = 0$.)

We think of the parameter $\lambda \geq 0$ 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.*

Assumption 1

(A1.1) $\quad r_{(0,N)}(0) = \ell_{(0,N)}(N) = 0;$

(A1.2) $\quad 0 \leq x < N \Rightarrow r_{(\lambda,N)}(x) > 0;$

(A1.3) $\quad 0 < x \leq N \Rightarrow \ell_{(\lambda,N)}(x) > 0;$

(A1.4) $\quad 0 < h \leq \frac{r_{(\lambda,N)}(x)}{\ell_{(\lambda,N)}(x)} \leq H < \infty;$

(A2.5) $\quad 0 < \lim_{\lambda\to 0} \frac{r_{(\lambda,N)}(0)}{\ell_{(\lambda,N)}(N)} < \infty.$

Assumption (A1.1) 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.[5] Assumptions (A1.2)–(A1.3) capture the requirement that mutations are always able to shift the population from any population state to a neighboring state. We shall shortly interpret Assumption (A1.4) as ensuring that our muddling agents are not too close to being maximizers. Assumption (A1.5) excludes a pathological case.

Assumption 2 *There exist functions $r_\lambda(k)$ and $\ell_\lambda(k)$ which are continuous for $0 \leq \lambda \leq 1$ and $0 \leq k \leq 1$ such that*

(A2.1) $\quad r_{(\lambda,N)}(kN) = r_\lambda(k) + O(\frac{1}{N});$

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

$$\text{(A2.2)} \quad \ell_{(\lambda,N)}(kN) = \ell_\lambda(k) + O(\tfrac{1}{N}).$$

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

Assumption 3

$$\text{(A3.1)} \quad 0 < k < 1 \Rightarrow (r_0(k) > 0 \text{ and } \ell_0(k) > 0)\,;$$

$$\text{(A3.2)} \quad \pi_X(k) > \pi_Y(k) \Longleftrightarrow r_0(k) > \ell_0(k);$$

$$\text{(A3.3)} \quad \pi_X(k) < \pi_Y(k) \Longleftrightarrow r_0(k) < \ell_0(k).$$

Assumption 3 is the essence of the muddling model. Assumption (A3.1) requires that some muddling is always present in the learning process (except in the pure population states $x = 0$ and $x = N$). Assumptions (A3.2)–(A3.3) require that the system is always more likely to move in the direction of a best reply than away from it. In light of this, Assumption (A1.4) 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.[6]

Aspirations and Imitation: an Example. This section presents a simple learning model satisfying Assumptions 1–3. Binmore, Samuelson and Vaughan [6] present a biological example.

In this example, an agent who receives the learn-draw recalls her average realized payoff in the last period and assesses it as being either "satisfactory" or "not satisfactory".[7] 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, labelling it unsatisfactory, and abandons the strategy. We refer to the probabilities that a

[6] Blume [7] 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.

[7] Satisficing models have long been the primary alternative to models of fully rational behavior, being pioneered in economics by Simon [28, 29, 30] and in psychology by Bush and Mosteller [8] (which remains a standard source on theories of learning), 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, Mookherjee and Ray [2] and Gilboa and Schmeidler [15, 17, 16].

player who has received the learn-draw will lose faith in and abandon her current strategy as *death probabilities* in order to stress the mathematical parallels between this model and that of Binmore, Samuelson and Vaughan [6]. For each expected payoff π, the corresponding death probability is given by

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

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.[8] The log-concavity of F 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.[9] 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.[10] For example, suppose that the rewards A and D each exceed B and C, so that the game has two 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.[11]

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

[8] This means that $\ln F$ is concave. See Bagnoli and Bergstrom [1] for a discussion of log-concavity and its implications. Many common distributions are log-concave, including the Chi, Chi-Squared, Exponential, Gamma, Logistic, Log Normal, Normal, Pareto, Poisson, Uniform, and Weibull distributions.

[9] 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.

[10] 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.

[11] 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.

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.

Payoffs and Learning. We have been working with a fixed specification of payoffs A, B, C and D and have accordingly been able to express transition probabilities as a function solely of the number of agents playing each strategy. Variations in the payoffs A, B, C and D will affect the learning process and hence transition probabilities, and we would like to derive the comparative static implications of these variations. To do so, we need to specify how the learning process depends upon payoffs. The aspiration and imitation model suggests the following assumptions:[12]

Assumption 4

(A4.1) $(A = D \ \ and \ \ B = C) \Rightarrow r_0(k) = \ell_0(1-k)$;

(A4.2) $r_0(k)$ *is increasing in* A *and* C, *decreasing in* B *and* D;

(A4.3) $\ell_0(k)$ *is increasing in* B *and* D, *decreasing in* A *and* C.

Assumption (A4.1) is a symmetry requirement, indicating that the learning process is driven by payoffs and attaches no particular importance to the strategy labels X and Y. The remaining assumptions imply that agents become more likely to switch to a strategy as its payoff increases and the payoff of the other strategy decreases.

Assumption 5

(A5.1) *Let* $\pi_X(h) > \pi_X(k)$. *Then* $r_0(h)r_0(k)/\ell_0(h)\ell_0(k)$ *increases when* A *is increased and* C *is decreased in such a manner that* $\pi_X(h)$ *is increased by the same amount that* $\pi_X(k)$ *is decreased.*

(A5.2) *Let* $\pi_Y(h) > \pi_Y(k)$. *Then* $\ell_0(h)\ell_0(k)/r_0(h)r_0(k)$ *increases when* D *is increased and* B *decreased in such a manner that* $\pi_Y(h)$ *is increased by the same amount that* $\pi_Y(k)$ *decreases.*

To interpret (A5.1), begin with the special case of $A = B = C = D$ and choose $h > \frac{1}{2}$ and $k = 1 - h < \frac{1}{2}$. Now let A increase and C decrease by a like amount. This causes $\pi_X(h)$ to increase and $\pi_X(k)$ to decrease by a like amount (given $k = 1-h < \frac{1}{2}$). Then, by (A4.2)–(A4.3), $r_0(h)/\ell_0(h)$ increases

[12]Increasing and decreasing are meant in their weak senses here.

and $r_0(k)/\ell_0(k)$ decreases. Assumption (A5.1) compares the elasticities of these two movements, asserting that the elasticity is higher for the increase in $r_0(h)/\ell_0(h)$, causing $r_0(h)r_0(k)/\ell_0(h)\ell_0(k)$ to increase. Assumption (A5.2) makes an analogous assertion for variations in B and D.

To verify that (A5) is satisfied by the aspiration and imitation model, we first calculate $r_{(\lambda,N)}(x)$. For the number of agents playing X to increase, given that an agent has received the learn-draw, three events must occur: (1) 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 $(N-x)/N$. (2) 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). (3) 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

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

Similarly,

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

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

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

Assumptions (A5.2) and (A5.3) now follow from differentiating (4) and using the facts that $g(\pi_X(k)) = F(\Delta - \pi_X(k))$, $g(\pi_Y(k)) = F(\Delta - \pi_Y(k))$, and F is log-concave.

3 Dynamics

Stationary Distribution. To examine the ultralong-run behavior of our learning model, we must study the stationary distribution of the system. In this section, we require only Assumptions 1–3.

Fix the game $\mathcal{G}$. 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 (A1.2)–(A1.3) ensure that for any state $x \in \{0,1,\dots N\}$, 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 given by the stationary distribution, hereafter simply called the "stationary distribution". Then $\gamma_{(\lambda,N,\tau)}(x)$ is the probability attached by the stationary distribution to state x. We study the stationary distribution $\gamma_{(\lambda,N)}$ obtained from $\gamma_{(\lambda,N,\tau)}$ by taking the limit $\tau \to 0$.

Working in the limit as $\tau \to 0$ ensures that the event in which more than one agent receives the learn-draw occurs with negligible probability. The model then becomes a birth–death process, as studied in Karlin and Taylor [20, ch. 4]. The following proposition makes this argument precise, where (5) is known as the "detailed balance" equation:

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:*

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

Proof.[13] Let $\Gamma^*_{(\lambda,N)}$ be the irreducible Markov chain whose transition matrix assigns the probabilities $r_{(\lambda,N)}(x)$, $1-r_{(\lambda,N)}(x)-\ell_{(\lambda,N)}(x)$, and $\ell_{(\lambda,N)}(x)$

[13] The techniques of Freidlin and Wentzell have become common, and can be used to give a alternative proof of this result. Freidlin and Wentzell ([12], Lemma 3.1 on page 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

to the events that the system moves from state x to states $x-1$, x, and $x+1$ respectively. This is the applicable transition matrix if ***exactly*** one player receives the learn-draw in each period. $\Gamma^*_{(\lambda,N)}$ is a birth–death process (which does not depend upon τ), and hence its stationary distribution must satisfy (5) by the standard theory.

Next, let $\Gamma^1_{(\lambda,N,\tau)}$ be the irreducible transition matrix contingent upon at least one (but possibly more than one) learn-draw being received in each period. Then because $\lim_{\tau\to 0}\Gamma^1_{(\lambda,N,\tau)} = \Gamma^*_{(\lambda,N)}$, the limiting stationary distribution of $\Gamma^1_{(\lambda,N,\tau)}$ as $\tau \to 0$ is the same as the stationary distribution of $\Gamma^*_{(\lambda,N)}$.

It thus suffices to show that $\Gamma_{(\lambda,N,\tau)}$ and $\Gamma^1_{(\lambda,N,\tau)}$ have the same limiting stationary distributions as $\tau \to 0$. Notice that

$$\Gamma_{(\lambda,N,\tau)} = (1-\tau)^N I + (1-(1-\tau)^N)\Gamma^1_{(\lambda,N,\tau)}, \tag{6}$$

where $(1-\tau)^N$ is the probability that no learn-draw is received by any player and I is the identity matrix. But (6) ensures that for every τ, $\Gamma_{(\lambda,N,\tau)}$ and $\Gamma^1_{(\lambda,N,\tau)}$ have the same stationary distribution and hence have the same limiting stationary distribution as $\tau \to 0$. □

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 λ is sufficiently small and N large, $\gamma_{(\lambda,N)}(x+1) > \gamma_{(\lambda,N)}(x)$ whenever $x > x^*$ (because strategy X must be a best reply here, and hence (A3.2) gives $r_0(x) > \ell_0(x)$). The stationary distribution $\gamma_{(\lambda,N)}$ must then increase on $[x^*, N]$. Similarly, from (A3.3), $\gamma_{(\lambda,N)}(x+1) < \gamma_{(\lambda,N)}(x)$, and $\gamma_{(\lambda,N)}(x)$ must decrease on $[0, x^*]$. The graph of γ therefore reaches maxima at 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 Figure 2.

(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 $x+1$ (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).

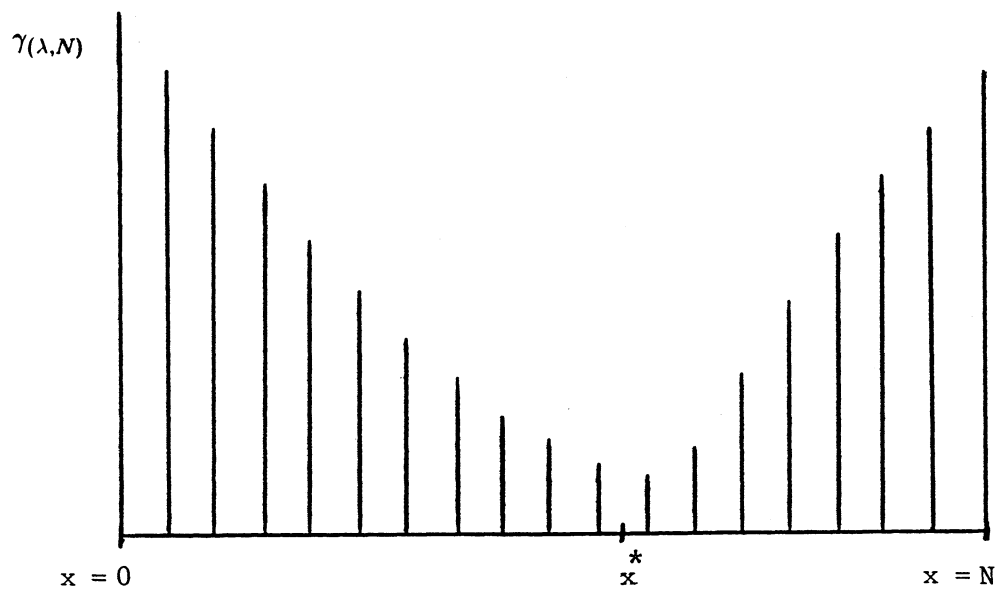

Figure 2: Stationary Distribution

Convergence. How long is the ultralong run? We provide a comparison of the convergence properties (for fixed N and very small mutation probabilities) of our muddling model and the model of Kandori, Mailath and Rob [19]. We consider the case of a game with two strict Nash equilibria.

Let $\Psi_{(\lambda,N)}$ be the transition matrix of the Kandori, Mailath and Rob model given mutation rate λ and population size N, and let $\psi_{(\lambda,N)}$ be its stationary distribution. Consider the following measure, which is examined by Ellison [10]:

$$\sup_{\psi^0} \limsup_{t\to\infty} \|\psi^0[\Psi_{(\lambda,N)}]^t - \psi_{(\lambda,N)}\|^{\frac{1}{t}},$$

where ψ^0 is the initial distribution. This is a measure of the distance between the distribution at time t (given by $\psi^0[\Psi_{(\lambda,N)}]^t$) and the stationary distribution (given by $\psi_{(\lambda,N)}$).

Kandori, Mailath and Rob [19] concentrate their attention on the limiting case of arbitrarily small mutation rates, perhaps reflecting a belief that changes in strategies are almost certainly driven by best-response considerations. Ellison shows that there exists a function $h_\psi: \mathbb{R} \to \mathbb{R}$ such that

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

where 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.[14]

We now seek an analogous measure of the rate at which our muddling model converges. Doing so requires that we address two issues. First, how is our continuous-time model to be compared to the discrete model of Kandori, Mailath and Rob?

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 choose the units in which time is measured so that the episodes in which every agent learns in the Kandori, Mailath and Rob model occur at each of the discrete times $1, 2, \ldots$. We then let $\beta = 1$, where β is the probability of a birth per unit 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, Mailath and Rob model.

The second problem concerns very small mutation rates. Mutation is the only source of noise in Kandori, Mailath and Rob, while our model retains noise in the learning process as $\lambda \to 0$. One might argue that a fair comparison of waiting times should require the noise in our learning process also become negligible, which would cause the waiting times for our process to increase tremendously. But this misses the point we want to make with the comparison. Given that 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.

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. $\Gamma_{(\lambda,N,\tau)}$ 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.

[14]We say that the functions $f(\lambda)$ and $g(\lambda)$ are comparable and write $f \sim g$, if there exist constants c and C such that for all sufficiently small λ, $c|g(\lambda)| \leq |f(\lambda)| \leq C|g(\lambda)|$.

Proposition 3 *There exists a function $h_\gamma(\lambda)$ such that*

$$\lim_{\tau\to 0}\left(1-\sup_{\gamma^0}\limsup_{t\to\infty}\|\gamma^0[\Gamma_{(\lambda,N,\tau)}]^{\frac{t}{\tau}}-\gamma_{(\lambda,N)}\|^{\frac{1}{t-1}}\right)\leq h_\gamma(\lambda)\sim\lambda, \tag{8}$$

where γ^0 is the initial distribution.

The proof is contained in the Appendix. Together, (7) and (8) imply that for very small values of λ, the muddling model converges much faster than does the Kandori, Mailath and Rob model. In particular, let $T_\psi(\eta)$ be the length of time required for the Kandori, Mailath and Rob model to be within η of its stationary distribution. Let $T_\gamma(\eta)$ be similarly defined for our muddling model. Then from (7) and (8), we have $\eta = (1-h_\psi(\lambda^z))^{T_\psi(\eta)}$ and $\eta \geq (1-h_\gamma(\lambda))^{T_\gamma(\eta)-1}$, giving, for small values of λ,

$$\frac{T_\psi(\eta)}{T_\gamma(\eta)-1}\geq\frac{\ln(1-h_\psi(\lambda^z))}{\ln(1-h_\gamma(\lambda))}\approx\frac{h_\gamma(\lambda)}{h_\psi(\lambda^z)}\sim\frac{1}{\lambda^{z-1}}. \tag{9}$$

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, Mailath and Rob 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^{33} ($>$ 8 billion) times as long to converge.

What drives this difference in rates of convergence? The Kandori, Mailath and Rob 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. Once a single mutation has allowed this escape, then the noisy learning dynamics can allow the system to "swim upstream" out of its basin of attraction.[15] The probability of 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 very 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. In [6], we present an example in which $N = 100$, $z = 33$, and $\lambda = .001$. The expected waiting time in the Kandori, Mailath and Rob model is approximately 1.7×10^{72}, while that of the muddling model is approximately 5000. But what about larger mutation rates? We expect the waiting times to be closer for larger mutation rates because increasing λ makes the Kandori, Mailath and Rob model noisier, reducing its waiting time.

We have examined waiting times for a fixed population size N. The next section shows that our model yields sharp equilibrium selection results as $N \to \infty$, but the model need not do so for small values of N. Does it then make sense to examine waiting times for fixed values of N? In many cases, 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.[16] On the other hand, if one is interested in large values of N, then an analogous argument to that behind Proposition 3 gives:

Proposition 4 *For sufficiently small* λ,

$$\lim_{N \to \infty} T_{\psi}(\eta)/T_{\gamma}(\eta) = \infty.$$

Finally, it is important to note that faster convergence rates still need not be fast enough. Convergence in our model is not fast in the second sense which Ellison [10] discusses. In particular, our waiting times do not remain bounded as N gets large. Hence, there remains plenty of room

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

[16] N often need not be very large before most of the mass of the stationary distribution is attached to a single state. For example in the example in [6], in which $N = 100$, $z = 33$, and $\lambda = .001$, the stationary distribution places more than .97 probability on states in which at most .05 of the population plays strategy X.

for skepticism as to the applicability of ultralong-run analyses based on examining stationary distributions.

4 Equilibrium Selection

We now consider equilibrium selection. To do this, we continue to examine the limiting stationary distribution of the Markov process as $\tau \to 0$, but now we also consider the limit as the population size gets large and the mutation rate gets small. In particular, we begin with the limiting stationary distribution 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. We again assume throughout this section that Assumptions 1–3 hold (but do not invoke Assumptions 4–5).

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 let $\gamma_{(\lambda,N)}$ denote the limiting stationary distribution of the Markov process on $\{0, 1, \ldots, N\}$ as $\tau \to 0$. Abusing notation, 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:

$$\int_0^1 (\ln r_0(k) - \ln \ell_0(k))dk \neq 0\,, \tag{10}$$

where (A1.4) and (A2.1)–(A2.2) ensure that the integral exists.

Proposition 5 *Let (10) 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 (A1.1)–(A1.3), we have

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

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 $x \in \{1, 2, \ldots, N-1\}$ (by (A1.4), 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:

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

and hence ((A1.5) ensuring that the limit exists)

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

Assumption 2 and (A1.4) ensure that the Riemann integral $\int_0^1(\ln r_0(k) - \ln \ell_0(k))dk$ exists, and hence we have:

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

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,[17] we then have $\lim_{N\to\infty} \lim_{\lambda\to 0} \gamma_{(\lambda,N)} = \gamma^*$.[18]

It remains to show 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 $[0,k_1]$ and $[k_2,k_3]$, for $0 < k_1 < k_2 < k_3 < k^*$, where k^* is the probability attached to strategy X by the mixed strategy equilibrium. Then

$$\frac{\gamma_{(\lambda,N)}([0,k_1])}{\gamma_{(\lambda,N)}([k_2,k_3])} \geq \prod_{x=k_1 N}^{k_2 N-1} \frac{r_{(\lambda,N)}(x)}{\ell_{(\lambda,N)}(x+1)}.$$

[17] 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 (10).

[18] This is a weak convergence claim. By Theorem 2.2 of Billingsley [3], 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).

For sufficiently small λ, (A3.2)–(A3.3) ensure that every term in the product on the right side of this inequality is less than one. Then for sufficiently small λ, $\lim_{N\to\infty} \gamma_{(\lambda,N)}([k_2, k_3]) = 0$. A similar argument applies to closed subintervals of $[k^*, 1]$ and yields the result.

Next, fix k arbitrarily small and consider the sets $[0, k)$ and $(1-k, 1]$.[19] We have

$$\frac{\gamma_{(\lambda,N)}([0,k))}{\gamma_{(\lambda,N)}((k,1])} = \prod_{x=kN}^{(1-k)N-1} \{\ln r_{(\lambda,N)}(x) - \ln \ell_{(\lambda,N)}(x)\}.$$

Then taking the limits $N \to \infty$ and $\lambda \to 0$ gives:

$$\lim_{\lambda\to 0} \lim_{N\to\infty} \frac{\gamma_{(\lambda,N)}([0,k))}{\gamma_{(\lambda,N)}((k,1])} = \int_k^{1-k} (\ln r_0(k) - \ln \ell_0(k))dk. \tag{12}$$

But for sufficiently small k, (12) takes the same sign as (11). 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 (10) 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, Mailath and Rob [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.

Proposition 6 (P6.1) *The selected equilibrium will be* (X, X) $[(Y, Y)]$ *if*

$$\int_0^1 \ln r_0(k)dk \;>\; [<] \;\int_0^1 \ln \ell_0(k))dk. \tag{13}$$

[19]We must consider $[0, k)$ and $(1-k,]$ rather than $\{0\}$ and $\{1\}$ because we do not know that $\lim_{N\to\infty} \gamma(\{0,1\}) = 1$

(P6.2) *The payoff-dominant equilibrium in game $\mathcal{G}$ can be selected even if it fails to be risk dominant.*

Proof. Item (P6.1) follows immediately from (11). To establish (P6.2), consider the aspiration and imitation model. Let

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

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 rewards, with

$$g(2) = 0 \quad g(1) = \tfrac{1}{4} \quad g(0) = \tfrac{1}{2} \quad g(-1) = \tfrac{3}{4}.$$

Inserting these probabilities in (2)-(3), taking the limits $\lambda \to 0$ and $N \to \infty$ and inserting in (11) gives:

$$\begin{aligned}
\frac{\gamma^*(1)}{\gamma^*(0)} &= \lim_{N\to\infty} N \int_0^1 (\ln g(\pi_Y(k)) - \ln g(\pi_X(k))dk \\
&= \lim_{N\to\infty} N \int_0^1 (\ln(\tfrac{1}{4}k + \tfrac{1}{2}(1-k)) - \ln(0k + \tfrac{3}{4}(1-k)))dk \\
&= \lim_{N\to\infty} \left(\frac{4}{3}\right)^N, \qquad (14)
\end{aligned}$$

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, Mailath and Rob [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, Mailath and Rob, 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 $k^*/N > 1/2$ (where k^* is the probability attached to X by the mixed-strategy equilibrium); 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 $\tilde{r}_0(k) = \phi B_X(k) + (1-\phi)r_0(k)$ and let $\tilde{\ell}_0(k) = \phi B_Y(k) + (1-\phi)\ell_0(k)$, 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. As ϕ increases to unity, the learning dynamics associated with $\tilde{r}$ and $\tilde{\ell}$ then approach best-reply dynamics.

Proposition 7 *For values of ϕ sufficiently close to one, the selected equilibrium is the risk-dominant equilibrium in game $\mathcal{G}$.*

Proof. Let $k^* \equiv x^*/N > 1/2$, so that (Y, Y) is the risk-dominant equilibrium. From (13), the selected equilibrium will be (Y, Y) if:

$$\begin{aligned}
&\int_0^{k'} \{\ln((1-\phi)r_0(k)) - \ln(\phi + (1-\phi)\ell_0(k))\}dk \\
&+ \int_{k'}^{k^*} \{\ln((1-\phi)r_0(k)) - \ln(\phi + (1-\phi)\ell_0(k))\}dk \\
&+ \int_{k^*}^{1} \{\ln(\phi + (1-\phi)r_0(k)) - \ln((1-\phi)\ell_0(k))\}dk < 0,
\end{aligned}$$

where $k' > 0$ satisfies $1 - k^* = k^* - k'$. Because $1 - k^* = k^* - k'$, the sum of the second and third terms on the left approaches a finite number as ϕ approaches unity. The first term approaches negative infinity, and hence the result. $\square$

Background Fitness. Another criterion for the risk dominant equilibrium to be selected emerges from considering the implications of background fitness. The concept of background fitness is borrowed from biology, where it refers to a process in which agents die and give birth randomly, regardless of payoffs. In particular, suppose that in each period, a randomly chosen agent dies and another randomly chosen agent gives birth to a daughter identical to herself. In the limiting case as $N \to \infty$ and $\lambda \to 0$, the probability of increasing or decreasing the number of agents playing X by one is then $k(1-k)$.[20]

Now suppose that the background fitness process governs strategy adjustments with probability θ, while with probability $1-\theta$ adjustments are given by $r_{(\lambda,N)}(x)$ and $\ell_{(\lambda,N)}(x)$, which satisfy Assumptions 1–3. Then we have the probabilities $\tilde{r}_{(\lambda,N)}$ and $\tilde{\ell}_{(\lambda,N)}$ given by:

$$\begin{aligned}\tilde{r}_0(k) &= \theta k(1-k) + (1-\theta) r_0(k) &(15)\\ \tilde{\ell}_0(k) &= \theta k(1-k) + (1-\theta)\ell_0(k). &(16)\end{aligned}$$

We might say that $\tilde{r}_0(k)$ and $\tilde{\ell}_0(k)$ have *background fitness* level θ in this case. If $\theta = 1$, then the process is driven *entirely* by background considerations, in that the evolution of the system has nothing to do with the payoffs in the game $\mathcal{G}$. If $\theta = 0$, then there are no background considerations and only payoffs in the game affect its evolution.

In our model, a background fitness level of θ will arise if, in each period of length τ, each agent receives a draw with probability $\theta\tau$ that causes that agent to pick another agent at random and imitate the strategy of the latter. If such a draw does not occur, then the agent's behavior is governed by the learning process. There are several alternative interpretations of the level of background fitness. For example, agents may be occasionally withdrawn from the game-playing population to be replaced by inexperienced novices. Alternatively, bounded rationality might be at work, there being some probability that an agent's thought processes are currently congested with other considerations. In our example, it may be that the agent is actually participating in a large number of similar games using the same strategy for each, as in the model of Carlsson and Van Damme [9]. Strategy revisions will then often be determined by what happens in games other than the game actually under study.

Proposition 8 *For sufficiently large levels of the background fitness level θ, the equilibrium (X,X) is selected if*

$$\int_0^1 \frac{1}{k(1-k)}(\ell_0(k) - r_0(k))dk < 0, \qquad (17)$$

[20]The probability of increasing the number of agents playing X by one is the probability that the agent who dies is playing strategy Y (given by $(N-x)/N = 1-k$ in the limit as N gets large) and the probability that the agent giving birth plays strategy X (given by $x/N = k$ for the limiting case of large N and small λ), giving $k(1-k)$.

and (Y, Y) is selected if (17) is positive.

Proof. Let (X, X) and (Y, Y) be Nash equilibria, with (Y, Y) being risk-dominant. From (13), the selected equilibrium will be (Y, Y) if

$$\int_0^1 \{\ln(k(1-k)\theta + (1-\theta)r_0(k)) - \ln(k(1-k)\theta + (1-\theta)\ell_0(k))\}dk < 0, \quad (18)$$

As $\theta \longrightarrow 1$, the left side of (18) approaches zero. We accordingly examine the first derivative of the left side of (18). A positive derivative ensures that (18) approaches zero from below, so that (18) is negative and hence the risk-dominant equilibrium is selected for θ near 1. Evaluated at $\theta = 1$, this derivative is given by the left side of (17). The selected equilibrium, for large background fitness, is thus (X, X) if (17) is negative and (Y, Y) if (17) is positive. □

What can we make of condition (17)? Because $r_0(k^*) = \ell_0(k^*)$ (from (A3.2)–(A3.3)), a straightforward evaluation of (17) gives:

Corollary 1 *Suppose $r_0(k)/k(1-k)$ and $\ell_0(k)/k(1-k)$ are linear in k. Then for sufficiently large θ, the risk-dominant equilibrium of game $\mathcal{G}$ is selected.*

We note that $r_0(k)/k(1-k)$ and $\ell_0(k)/k(1-k)$ are linear in k in the aspiration and imitation model when F is the uniform distribution.

We thus have another condition for the risk-dominant equilibrium to be selected. The condition comes in two parts. First, the learning process must be related to the proportions of agents playing the two strategies in a linear way, as in the aspiration and imitation model. Second, changes in strategy must be driven primarily by background fitness considerations and not payoffs in the game. Hence, we select the risk-dominant equilibrium in relatively unimportant games—those whose payoffs have little to do with agents' behavior.[21]

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,

[21] Many variations on Proposition 8, involving various rules under which transitions are given by the learning process with probability $1-\theta$ and with probability θ are determined by a random process exogenous to the game, and yielding the result that the risk-dominant equilibrium is selected if θ is sufficiently large and (17) holds, can be similarly established.

mixed-strategy Nash equilibrium. Then an argument analogous to that of Proposition 5 yields

Proposition 9 *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 long run outcome is the mixed strategy equilibrium. Our inclination here is to regard the latter as the more useful model. Binmore, Samuelson and Vaughan [6] discuss this issue at greater length.

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.

It is immediately apparent that 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 now invoke Assumptions 4–5 as well as Assumptions 1–3. We investigate games with two strict Nash equilibria ($A > B$ and $D > C$) and ask when the risk-dominant equilibrium will be selected.

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

Proposition 10 *If the payoff-dominant equilibrium in game $\mathcal{G}$ is also risk dominant, then the payoff-dominant equilibrium is selected.*

Proof: Let (X, X) and (Y, Y) be risk-equivalent in game $\mathcal{G}$, so that $A + C = B + D$, and let $A = D$. Then (A4.1) ensures that (13) holds with equality. Now make (X, X) the payoff-dominant equilibrium by increasing A and decreasing C so as to preserve $A+C = B+D$ (and hence to preserve

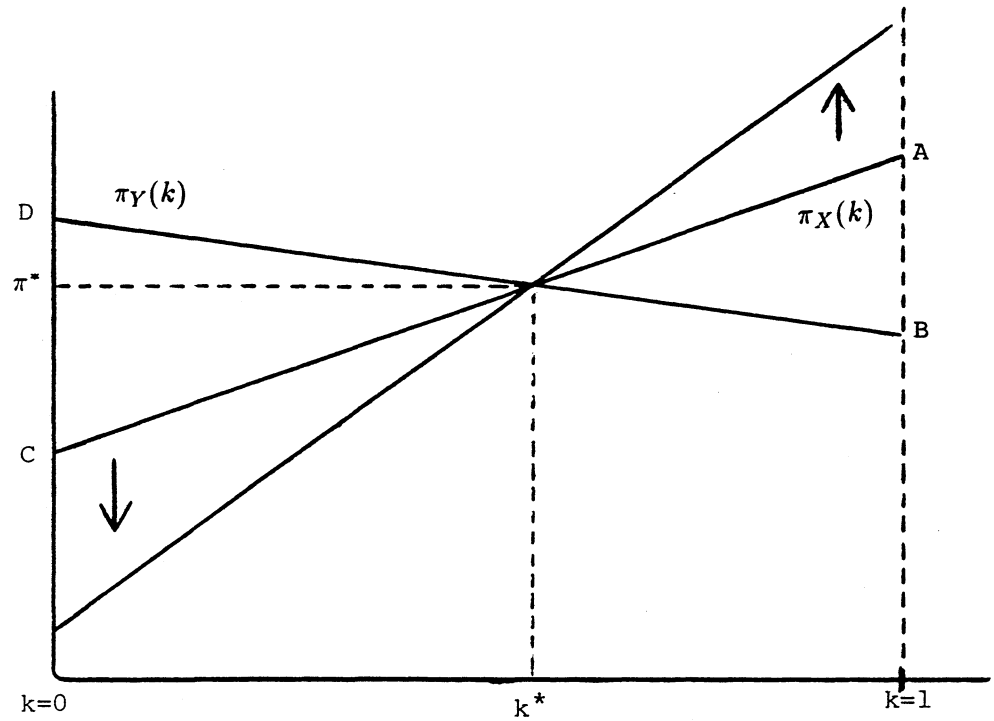

Figure 3: Payoff Variations

the risk-equivalence of the two strategies). Assumption (5) then ensures that $\int_0^1 \ln r_0(k) - \ell_0(k))dk$ increases. The payoff-dominant equilibrium (X, X) is then 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 at least weakly increases the function $\ln \pi_Y(k) - \ln \pi_X(k)$ on $[0, 1]$, which can only strengthen the inequality in (13) (by (A4.2)–(A4.3)), and hence preserves the result that the payoff-dominant equilibrium is selected. $\square$

We now consider cases where the payoff- and risk-dominance criteria conflict. We ask how the likelihood of choosing the risk-dominant equilibrium in the game $\mathcal{G}$ varies with the payoffs in the game. To pose this question precisely, let k^* be the probability attached to X in the mixed strategy equilibrium. Let (Y, Y) be the risk-dominant equilibrium, so $k^* > 1/2$, but let (X, X) be payoff-dominant. Let π^* be the payoff in game $\mathcal{G}$ 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 an increase in A accompanied by a decrease in C calculated so as to preserve k^* and π^*, as illustrated in Figure 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 $(1 - k^*)C(A) + k^*A =$

$(1-k^*)B(D)+k^*D=\pi^*$. Let $\Omega(k^*,\pi^*)=\{(A,D):D\in(C(A),A)\}$; this is the set of possible values of A and D that are consistent with (X,X) being the payoff-dominant equilibrium given the fixed values of k^* and π^*. Let $\Omega^*(k^*,\pi^*)\subset\Omega(k^*,\pi^*)$ be the subset of values for which the payoff-dominant equilibrium in the game $\mathcal{G}$ is selected. Then we have:

Proposition 11 (P11.1) *If $\Omega^*(k^*,\pi^*)$ contains (A,D) with $D>B(D)$, then $\Omega^*(k^*,\pi^*)$ contains every convex combination of (A,D) and (A,π^*).*

(P11.2) *Fix π^*. Then there exist values of k^* for which $\Omega^*(k^*,\pi^*)$ is nonempty, and if $\Omega^*(k^{**},\pi^*)$ is nonempty then $\Omega^*(k^*,\pi^*)$ is nonempty for all $k^*\in(\frac{1}{2},k^{**})$.*

Proposition (P11.1) tells us that if $D>B$ and the payoff-dominant equilibrium is selected, then the payoff-dominant equilibrium continues to be selected as D is decreased (and B increased), at least until $D=\pi^*$. Hence, movements in the payoffs to strategy Y that increase B and decrease D make the payoff-dominant equilibrium more likely to be selected. The best case for the payoff-dominant equilibrium occurs when $D=B$ or $D<B$. The payoff-dominant equilibrium is thus 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. Proposition (P11.2) indicates that the larger is the basin of attraction of the risk-dominant equilibrium, the harder it is for the payoff-dominant equilibrium to be selected.

Proof of Proposition 11: (P11.1) Fix $k^*>1/2$. Fix A and hence $C(A)$. If we set $D=A$, then Proposition 10 ensures that (Y,Y) will be selected, since it is risk dominant and payoff undominated. Now let D decline. From (A5.1)–(A5.1), $\int_0^1(\ln r_0(k)-\ell_0(k))dk$ declines until D reaches π^*. Hence, from (13), if the payoff-dominant equilibrium is selected for any value of D, then it is also selected for any smaller value $D\geq\pi^*$.

(P11.2) Let values A, B, C, and D exist such that the payoff-dominant equilibrium is selected and the mixed-strategy equilibrium is given by k^{**}, with payoff π^*. Let $1/2<k^*<k^{**}$. Then we can find values $A'>A$, $C'>C$, $B'<B$, and $D'<D$ that (1) give the mixed-strategy equilibrium k^*; (2) increase all payoffs to X and decrease all payoffs to Y; and (3) preserve π^*. From (A4.2)–(A4.3), these payoffs must then preserve the property that the payoff-dominant equilibrium is selected, so $\Omega^*(k^*,\pi^*)$ is nonempty. □

Some Experimental Evidence. Given Proposition 11, 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 altering a game in which $D < B$ to achieve $D = B$, while preserving the basins of attraction of the Nash equilibria, causes the risk-dominant equilibrium to be selected much more often and the payoff-dominant equilibrium to be selected much less often. Both experiments are consistent with our results.

6 Endogenous Learning

The heart of our model is a learning rule. If we want to sharpen our results by working with a specific learning rule, which rules are worthy of our attention? A useful way to approach this question is perhaps 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? Two considerations arise. First, the higher processes may well be biological and hence hard-wired into our cognitive apparatus. The second is that agents will learn only when there is something to learn about. We can therefore hope to escape the infinite regress by showing that outcomes are not particularly sensitive to the nature of the learning rules that agents use in learning how to learn. In particular, we look for cases in which learning rules satisfy conditions analogous to evolutionary stability.[22] Such rules will fare well in any process that rewards rules which generate above-average payoffs. 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 restrict attention to the aspiration and imitation model, and allow only the value of the aspiration level Δ to be subject to change.[23] 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 evolution of the aspiration level Δ depends upon the distribution of $\tilde{R}$, and especially on its dispersion.

Proposition 12 *Fix $z > 0$. Suppose that* $\text{prob}(R < \Delta - z)/\text{prob}(R < \Delta)$ *is increasing in Δ. Then for large enough N and small enough λ and for any Δ' with $\Delta' < \Delta$, the payoff to a player characterized by Δ' in any population consisting of these two rules exceeds that of Δ.*

To interpret Proposition 12, suppose that $\text{prob}(R < \Delta - z)/\text{prob}(R < \Delta)$ is increasing in Δ. This will hold for distributions that are not too dispersed, in the sense that the probability in the tails of the distribution falls off sufficiently rapidly as one moves out along the tail. For example, this condition holds for the Normal distribution. To see this, we note that, for the standard Normal, the condition is equivalent to the statement that $\text{prob}(R > s + z)/\text{prob}(R > s)$ is decreasing in s. Since

$$\frac{d}{ds}\left(\frac{\text{prob}(R > s + z)}{\text{prob}(R > s)}\right) = \frac{-e^{-(s+z)^2/2}\text{prob}(R > s) + e^{-s^2/2}\text{prob}(R > s + z)}{(\text{prob}(R > s))^2},$$

it suffices to show that

$$\frac{\text{prob}(R > s + z)}{\text{prob}(R > s)} < e^{-sz}e^{-z^2/2}.$$

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

[23]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, Mookerjee and Ray [2] suggest that aspiration levels should adjust to equal the average equilibrium payoff. In Binmore and Samuelson [5], we show that this is not always possible in a muddling model.

To establish this inequality, we note that $\text{prob}(R > s + z) =$

$$\int_{s+z}^{\infty} e^{-x^2/2}dx = \int_{s}^{\infty} e^{-(z+x)^2/2}dx = e^{-z^2/2}\int_{s}^{\infty} e^{-zx}e^{-x^2/2}dx$$

$$\leq e^{-z^2/2}e^{-sz}\int_{s}^{\infty} e^{-x^2/2}dx = e^{-z^2/2}e^{-sz}\text{prob}(R > s). \tag{19}$$

Proof of Proposition 12. 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 close to 0 and 1, 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 a 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

$$\Pi(P_Y) = \sum_{k=0}^{\infty}\left(\frac{kp_Y^k(N,\lambda)}{\sum_{h=0}^{\infty} hp_Y^h(N,\lambda)}\right)\Pi^k(P_Y)\,, \tag{20}$$

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

$$\lim_{N\to\infty}\sum_{h=0}^{\infty} hp_Y^h(N,\lambda) = \infty\,, \tag{21}$$

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 $\epsilon > 0$, there is a $\delta' > 0$ such that for all $\delta > \delta'$,

$$\Pi^{\delta}(P_Y) > \Pi^{\infty}(P_Y) - \epsilon . \tag{22}$$

Then (22) and (21) (along with $\Pi^{\infty}(P_Y) > 0$) imply the desired result that (20) 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 $\text{prob}(R < \Delta - z)/\text{prob}(R < \Delta)$ is increasing in Δ. In particular, for every state in the set P_Y, Y is a best reply and X is an inferior reply; and P_Y can be made sufficiently small that the lowest payoff to a best reply over states in this set exceeds the highest payoff to an inferior reply. 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 $\text{prob}(R < \Delta - z)/\text{prob}(R < \Delta)$ is increasing in Δ then the ratio of the probability of abandoning a best for an inferior reply to the probability of abandoning an inferior for a best reply is lower for aspiration level Δ' than 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. The former must then receive a higher payoff, yielding the result. $\square$

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 make them relatively unlikely to switch away from high payoff realizations and relatively likely to switch away from low payoff realizations. In the case of distributions like the Normal, for which $\text{prob}(R < \Delta - z)/\text{prob}(R < \Delta)$ is increasing in Δ, these learning rules involve smaller aspiration levels.

If $\text{prob}(R < \Delta - z)/\text{prob}(R < \Delta)$ is increasing in Δ, then the dynamics concerning aspiration levels are straightforward. Let the aspiration level initially be given by Δ. Let the population be sufficiently large and mutation rate sufficiently small. Then if an aspiration level $\Delta' < \Delta$ appears, it will displace Δ. This can in turn be displaced by a lower aspiration level Δ''. Continuing in this fashion, the process will push the aspiration level ever lower.[24] Whereas we have presented this result in terms of only two coexisting aspiration levels. If there are more than two aspiration levels, then the highest such levels will always earn lower payoffs than at least some lower levels, creating a downward pressure on aspiration levels.

What are the implications of this process for the selected equilibrium? Here we specialize to the standard Normal distribution to minimize notation.

Proposition 13 *Let F be the standard Normal distribution. Then for sufficiently small Δ, the selected equilibrium is the equilibrium that 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 (13)) that the risk-dominant equilibrium will be selected if $\ln F(\Delta - B) + \ln F(\Delta - D) - \ln F(\Delta - A) - \ln F(\Delta - C) < 0$. 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

[24] Fortunately, we do not have to worry about the possibility that lower values of Δ will vitiate the assumption N is sufficiently large and λ sufficiently small. Decreasing Δ increases the pressure towards the ends of the state space, ensuring that N will still be large enough and λ small enough to yield a stationary distribution sufficiently concentrated near the ends of the state space.

hold for large $-\Delta$ because (Y, Y) is risk-dominant, so that $A+C-B-D<0$ and hence the coefficient on $-\Delta$ is negative. $\square$

We therefore have conditions under which evolution leads to learning rules that will select the risk-dominant equilibrium.[25] However, we have examined the evolution of a very narrow class of learning rules, allowing only the aspiration level in the aspiration and imitation model to adjust. 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, in that we have examined only a very narrow class of learning rules. But we believe this to be an important area for further work. Another important area for further work is the extension of the analysis to larger games. The convenience of the detailed balance equation (5) will then be lost but we are hopeful that other techniques can be applied.

[25] Note that there are widely dispersed distributions, such as the Cauchy distribution, for which $\text{prob}(R+\Delta-z)/\text{prob}(R<\Delta)$ does not increase in Δ, and hence which do not push Δ ever lower, possibly allowing the payoff-dominant equilibrium may survive.

8 Appendix: 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.

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:

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

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 (8) to show that for any γ^0,

$$\lim_{\tau\to 0}\left(1-\|\gamma^0\sum_{k=0}^{t/\tau}\iota(k,\frac{t}{\tau})[\Gamma^*_{(\lambda,N)}]^k-\gamma_{(\lambda,N,\tau)}\|^{\frac{1}{t-1}}\right)\leq h_\gamma(\lambda)\sim\lambda. \tag{23}$$

Let $\tau_n = 1/n$. Then $t/\tau_n = nt$, an equality we shall use repeatedly. For each n, let $\{Z_{nk}, k \in \{1,\ldots,nt\}\}$ be a collection 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 $\{Z_{nh}, h\in\{1,\ldots,nt\}\}$ take the value one. Notice that, for any n, we have $\sum_{k=1}^{nt}\tau_n N = nt\tau_n N = tN$, so that for any n, the sum over the collection $\{Z_{nk}, k\in\{1,\ldots,nt\}\}$ of the probabilities of receiving the outcome one is finite and given by tN. Coupled with the fact that τ_n and hence $\tau_n N$ approach zero as n gets large, this allows us to apply Theorem 23.2 of Billingsley [4] to conclude that

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

Hence, as τ_n gets small, $\iota(k,\frac{t}{\tau_n}) = \iota(k,nt)$ is given by a Poisson distribution with mean and variance Nt. It accordingly suffices for (23) to show, for any γ^0, that

$$\|\gamma^0 \sum_{k=0}^{\infty} \frac{(Nt)^k}{e^{-Nt}k!}[\Gamma^*_{(\lambda,N)}]^k - \gamma_{(\lambda,N)}\| \leq (1 - h_\gamma(\lambda))^{t-1}. \tag{24}$$

We first observe that $\gamma_{(\lambda,N,\tau)}\hat{\Gamma}_{(\lambda,N,\tau)} = \gamma_{(\lambda,N,\tau)}(1 - \tau N)I + \tau N\Gamma^*_{(\lambda,N)}) = \gamma_{(\lambda,N,\tau)}$, which we can solve for $\gamma_{(\lambda,N)}\Gamma^*_{(\lambda,N)} = \gamma_{(\lambda,N)}I = \gamma_{(\lambda,N)}$. 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

$$\|\gamma^0[\Gamma^*_{(\lambda,N)}]^N)^t - \gamma_{(\lambda,N)}\| \leq (1 - 2S(\lambda))^{t-1}, \tag{25}$$

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 what the least likely transition in the matrix $[\Gamma^*_{(\lambda,N)}]^N$ is. 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 $S_M(\lambda)$ is given by

$$\prod_{x=0}^{N-1} r_{(\lambda,N)}(x, N) = \lambda[c' + h(\lambda)], \tag{26}$$

where c' does not depend on λ and $\lim_{\lambda\to 0} h(\lambda) = 0$, and where the equality follows because for all $x \in \{1, 2, \ldots, N-1\}$, $\lim_{\lambda\to 0} r_{(\lambda,N)} = r_{(0,N)} > 0$). Hence, for sufficiently small λ, we have $|h(\lambda)| < \epsilon$ for some $\epsilon > 0$. We then let $c = c' - \epsilon$ and $C = c' + \epsilon$ 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 (25) gives, for any $\alpha > 0$,

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

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

$$\|\gamma^0 \sum_{k=0}^{\infty} \frac{(Nt)^k}{e^{-Nt}k!}[\Gamma^*(\lambda, N)]^k - \gamma_{(\lambda,N)}\| \leq (1 - h_\gamma(\lambda))^{\alpha t-1} \tag{27}$$

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

$$\lim_{t\to\infty} \frac{(Nt)^k}{e^{-Nt}k!}[\Gamma^*_{(\lambda,N)}]^k = 0 \tag{28}$$

for any $k < \alpha Nt$. This in turn follows from noting that as t grows, the Poisson distribution with mean and variance Nt approaches a Normal distribution with mean and variance Nt (cf. Billingsley [4], problem 27.3 on page 379). Equation (28) is then $\lim_{t\to\infty} \text{prob}[\text{N}(Nt, Nt) < \alpha Nt] = \lim_{t\to\infty} \text{prob}[\text{N}(0,1) < (\alpha - 1)(Nt)^{\frac{1}{2}}] = 0$, which follows from the fact that $(\alpha - 1)(NT)^{\frac{1}{2}} \to \infty$ as $t \to \infty$.

Finally, we note that because (27) holds for any $\alpha \in (0,1)$, we must have (24), which is the desired result. □

References

[1] Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. Mimeo, University of Michigan, 1989.

[2] Jonathan Bendor, Dilip Mookherjee, and Debraj Ray. Aspiration–based adaptive learning in two person repeated games. Mimeo, Indian Statistical Institute, 1991.

[3] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, New York, 1968.

[4] Patrick Billingsley. *Probability and Measure*. John Wiley and Sons, New York, 1986.

[5] Ken Binmore and Larry Samuelson. Muddling through: Noisy equilibrium selection. Ssri working paper 9410, University of Wisconsin, 1993.

[6] Ken Binmore, Larry Samuelson, and Richard Vaughan. Musical chairs: Modelling noisy evolution. Mimeo, University College London and University of Wisconsin, 1993.

[7] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[8] Robert R. Bush and Frederick Mosteller. *Stochastic Models for Learning*. John Wiley and Sons, New York, 1955.

[9] Hans Carlsson and Eric van Damme. Global games and equilibrium selection. *Econometrica*, 61:989–1018, 1993.

[10] Glenn Ellison. Learning, local interaction, and coordination. Mimeo, MIT, 1992.

[11] Glenn Ellsion and Drew Fudenberg. Rules of thumb for social learning. IDEI discussion paper 17, University of Toulouse, 1992.

[12] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems.* Springer–Verlag, New York, 1984.

[13] Daniel Friedman. Equilibrium in evolutionary games: Some experimental results. Mimeo, University of California, Santa Cruz, 1993.

[14] Drew Fudenberg and Chris Harris. Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57:420–441, 1992.

[15] Itzhak Gilboa and David Schmeidler. Case–based decision theory. Mimeo, Northwestern University, 1992.

[16] Itzhak Gilboa and David Schmeidler. Case–based consumer theory. Mimeo, Northwestern University, 1993.

[17] Itzhak Gilboa and David Schmeidler. Case–based optimization. Mimeo, Northwestern University, 1993.

[18] C. B. Harley. Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, 89:611–633, 1981.

[19] Michihiro Kandori, George J. Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56, 1993.

[20] Samuel Karlin and Howard M. Taylor. *A First Course in Stochastic Processes.* Academic Press, New York, 1975.

[21] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains.* D. Van Nostrand Company, Inc., Princeton, 1960.

[22] John Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, Cambridge, 1982.

[23] Paul R. Milgrom. Good news and bad news: Representation theorems aned applications. *Bell Journal of Economics*, 12:380–391, 1981.

[24] R. Nelson and S. Winter. *An Evolutionary Theory of Economic Change.* Harvard University Press, Cambridge, 1982.

[25] Georg Nöldeke and Larry Samuelson. An evolutionary analysis of backward and forward induction. *Games and Economic Behavior*, 5:425–454, 1993.

[26] Arthur J. Robson. The biological basis of expected utility, knightian uncertainty and the ellsberg paradox. Mimeo, University of Western Ontario, 1992.

[27] Arthur J. Robson and Fernando Vega-Redondo. Efficient equilibrium selection in evolutionary games with random matching. Mimeo, University of Western Ontario and Universidad de Alicante, 1994.

[28] Herbert Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69:99–118, 1955.

[29] Herbert Simon. *Models of Man.* New York, 1957.

[30] Herbert Simon. Theories of decision making in economics and behavioral science. *American Economic Review*, 49:253–283, 1959.

[31] Paul G. Straub. Risk dominance and coordination failures in static games. Dispute Resolution Research Center Working Paper 106, Northwestern University, 1993.

[32] S. Winter. Satisficing, selection and the innovating remnant. *Quarterly Journal of Economics*, 85:237–260, 1971.

[33] Peyton Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.

RECENT SSRI WORKING PAPERS

This Series includes selected publications of members of SSRI and others working in close association with the Institute. Inquiries concerning publications should be addressed to Linda Strohmenger, SSRI, 6470 Social Science Bldg., University of Wisconsin-Madison, Madison, WI 53706. (608) 263-3876. Electronic mail address: SSRI@macc.wisc.edu. Gopher and anonymous FTP: gopher.econ.wisc.edu

9435
Jones, Larry E. and Rodolfo E. Manuelli
GROWTH AND THE EFFECTS OF INFLATION

9436
Staiger, Robert W. and Frank A. Wolak
DIFFERENCES IN THE USES AND EFFECTS OF ANTIDUMPING LAW ACROSS IMPORT SOURCES

9437
Fung, K.C. and Robert W. Staiger
TRADE LIBERALIZATION AND TRADE ADJUSTMENT ASSISTANCE

9438
George J. Mailath, Larry Samuelson and Jeroen M. Swinkels
HOW PROPER IS SEQUENTIAL EQUILIBRIUM?

9439
Raymond J. Deneckere and Dan Kovenock
CAPACITY-CONSTRAINED PRICE COMPETITION WHEN UNIT COSTS DIFFER

9440
Beth Allen,Raymond Deneckere,Tom Faith and Dan Kovenock
CAPACITY PRECOMMITMENT AS A BARRIER TO ENTRY: A BERTRAND-EDGEWORTH APPROACH

9441
SSRI TRIENNIAL REPORT
1991-1994

9442
Wen-Ling Lin
INTERDEPENDENCE BETWEEN INTERNATIONAL STOCK RETURNS -- NEWS OR CONTAGION EFFECT?

9443
Steven N. Durlauf
A THEORY OF PERSISTENT INCOME INEQUALITY

9444
Jeff Dominitz,Charles F. Manski
USING EXPECTATIONS DATA TO STUDY SUBJECTIVE INCOME EXPECTATIONS

9445
Blake LeBaron
TECHNICAL TRADING RULE PROFITABILITYAND FOREIGN EXCHANGE INTERVENTION

9446
Blake LeBaron
CHAOS AND NONLINEAR FORECASTABILITY IN ECONOMICS AND FINANCE

9447
Blake LeBaron and Andreas S. Weigend
EVALUATING NEURAL NETWORK PREDICTORS BY BOOTSTRAPPING

9448
Maria E. Muniagurria and Nirvikar Singh
FOREIGN TECHNOLOGY, SPILLOVERSAND R&D POLICY

9449
John Rust
NUMERICAL DYNAMIC PROGRAMMING IN ECONOMICS

9450
Jeff Dominitz and Charles F. Manski
ELICITING STUDENT EXPECTATIONSOF THE RETURNS TO SCHOOLING

9451
Robert W. Staiger
INTERNATIONAL RULES AND INSTITUTIONS FOR TRADE POLICY

9452R
Yeon-Koo Che and Ian Gale
AUCTIONS WITH FINANCIALLY-CONSTRAINED BIDDERS

9501
Yeon-Koo Che and Dietrich Earnhart
OPTIMAL USE OF INFORMATION IN LITIGATION: SHOULD REGULATORY INFORMATION BE WITHHELD TO DETER FRIVOLOUS SUITS?

9502
Arthur S. Goldberger and Charles F. Manski
REVIEW ARTICLE: THE BELL CURVE BY HERRNSTEIN AND MURRAY

9503
Carol L. Baker and Wen-Ling Lin
CREDIBILITY OF MEXICO'S CRAWLING PEG POLICY -- EVIDENCE FROM EXPECTATIONS PANEL DATA

9504
Kyle Bagwell and Robert W. Staiger
COLLUSION OVER THE BUSINESS CYCLE

9505
Charles F. Manski
LEARNING ABOUT SOCIAL PROGRAMS FROM EXPERIMENTS WITH RANDOM ASSIGNMENT OF TREATMENTS

9410R
Ken Binmore and Larry Samuelson
MUDDLING THROUGH: NOISY EQUILIBRIUM SELECTION

SOCIAL SYSTEMS RESEARCH INSTITUTE
University of Wisconsin Madison
1180 Observatory Drive
Madison, WI 53706 USA

WORKING PAPERS AND REPRINT ORDER FORM

Date: ______________________________

Name:______________________________

Address:____________________________

____

____

Working Paper or Reprint Number	Author or Title

RATES (Includes postage)

	U.S., Canada, Mexico	International
Working Paper or Reprint	$ 6.00	$ 8.00
Subscription	$ 60.00	$80.00

AMOUNT DUE:______________________________

AMOUNT ENCLOSED:_________________________

PLEASE RETURN THIS FORM WITH YOUR REMITTANCE PAYABLE TO SSRI