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

October 26, 1994

[1]First draft August 26, 1992. Financial support from the National Science Foundation and the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, 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 and Richard Vaughan for helpful discussions.

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 payoff-dominant equilibrium in a 2×2 game is selected by the model as well as conditions under which the risk-dominant equilibrium is selected. The relevant risk-dominance considerations, however, arise not in the original game but in a "fitness game" derived from the process by which payoffs in the original game are translated into evolutionary fitnesses. We also find that waiting times until the limiting distribution is reached can be shorter than in a mutation-driven model. To explore the robustness of the results to the specification of the model, we present a number of 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.

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 equilibriating process or "libration" is explicitly modeled. A boundedly rational player is identified with a learning rule and attention is centered on the interactive dynamics that result when pairs of players are repeatedly drawn at random from a given population to play a given game.

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, the learning rules studied in a dynamic treatment may seem overly-specific or even arbitrary, and 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 to criticize a dynamic model for such reasons is to miss an important point. 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.

At the same time, dynamic models are unlikely to yield much insight into which aspects of a game's environment are significant in equilibrium selection if the mathematical properties of the learning rules studied are simply plucked from the air. For this reason, we consider the

question of **microfoundations** to be crucial: it is important to derive the model's equations of motion from explicitly stated assumptions about the manner in which individual learning is postulated to proceed. To facilitate this task in

the current paper, a study of the more technical issues has been relegated to Binmore, Samuelson and Vaughan [**?**], in which the model is motivated by a story of a competition for survival between stylized rabbits. The present paper provides socio-economic microfoundations for the same model and uses the model to explore the following issues:

(a) Time: We have distinguished elsewhere ([**?**, **?**]) between four relevant time periods: the short run, the medium run, the long run, and the ultralong run. In the short run, the system will have little chance to stray from its initial condition. In the medium run, the system will begin to respond to evolutionary pressures. In the long run, the system may find its way to an equilibrium, with this equilibrium being determined by the initial conditions of the system. Only in the ultralong run, history can be neglected for the purposes of equilibrium selection. In the ultralong run, random shocks will repeatedly bounce the system out of the basin of attraction of one equilibrium into the basin of attraction of another. The system may then eventually settle into a steady state in which each equilibrium is visited with a well-defined frequency. If we are lucky, all but one of these frequencies may be negligible. If so, then an equilibrium has been selected in the ultralong run.[1]

The appropriate technique of analysis depends upon the time period of interest.

Our concern in this paper is with equilibrium selection in the ultralong run. We are aware that the interest of such results depends on the waiting time before we can expect an ultralong-run prediction to be realized. Such waiting times depend upon the manner in which noise enters the model, prompting the next issue.

(b) Noise: For ultralong-run phenomena to be meaningful, it is necessary to model the noise that perturbs the learning process and drives the selection of equilibria in the ultra-long run. The pioneers in this regard are Young [**?**] and Kandori, Mailath and Rob [**?**]. In their models, agents normally choose a best

response given their information, and hence act 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 $\lambda \to 0$.

[1] Kandori, Mailath and Rob [**?**] call such an equilibrium a "long-run equilibrium", while for us it is an "ultralong- run equilibrium".

We proceed in the same manner except that our agents are *muddlers* rather than maximizers. For muddlers, the learning process is itself noisy, in that agents do not always choose best responses. We think it necessary to model such "selection noise" explicitly when mutation rates are allowed to

become very small because selection noise will then be the major source of randomness in the system. Such an innovation has the important 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, in the long run, to an equilibrium that is not selected in the ultralong run. In the maximizing models of Young [**?**] and Kandori, Mailath and Rob [**?**], a large number of *simultaneous* mutations are now necessary for the system to escape from its 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.

Incorporating noisy learning into the model has implications for equilibrium selection as well as waiting times: 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.

(c) Payoffs: In conventional game theory, the payoffs are Von Neumann and Morgenstern utilities. In economic applications, the players are often assumed to be risk-neutral so that their utilities can be identified with physical *rewards*. For example, firms are usually assumed to maximize expected profit. In contrast, payoffs in biological models are taken to be *fitnesses*, generally measured in units like the expected number of offspring, rather than physical rewards (such as food or square meters of territory). When maximizing models are used, the distinction between a fitness and a reward is seldom important. In our muddling model, however, we find it essential to translate rewards into fitnesses and to work with the latter. For example, we find that the relevant risk-dominance

considerations in our model require the use of fitnesses rather than rewards. Except in special cases, there is no straightforward relation between risk-dominance

in the reward game and in the fitness game. Even in cases where risk-dominance considerations in fitnesses and rewards coincide, one must beware of treating the payoffs in the fitness game like Von Neumann and Morgenstern payoffs. In particular, equilibrium selection in a muddling model is not invariant to strictly affine transformations of the fitnesses.

(d) Robustness Why study our model of muddled learning rather than one of the many alternatives that might be proposed? As in most studies, our choice of

model was constrained by the need to keep the mathematics simple and by a desire for our results to be comparable with those obtained by others.[2] Operating under such constraints, one is inevitably led to make questionable modeling assumptions. However, we hope ultimately to dispense with the need to

make arbitrary choices in the construction of the model by treating the learning process itself as being determined by evolutionary processes.

To illustrate this proposed methodology, we conclude this paper with an examination of the evolutionary stability of the learning rules within the narrow class of rules considered in this paper. 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 class. If it can, then we have grounds for questioning its robustness. If it cannot be invaded, then we say that the learning rule is itself evolutionarily stable. We find conditions under which evolutionarily stable learning rules in our muddling model select the risk-dominant equilibrium in the reward game for symmetric 2×2 games, thus matching the results of maximizing models. In other cases, however, the models continue to give differing results.

The paper is organized as follows. Section 2 describes the muddling model and introduces the distinction between a reward and a fitness. Section 3 reviews 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. Section 5 is devoted to comparative statics. Section 6 considers the evolutionary stability of the learning rules studied. Section 7 briefly discusses our conclusions.

[2] For example, Binmore, Samuelson and Vaughan [**?**], studying the long run, show that the muddling model is approximated by the replicator dynamics when only long-run considerations are relevant.

2 A Muddling Model

The Reward Game. We begin with the symmetric 2×2 game $\mathcal{R}$ of Figure 1. The payoffs in this game, taken to be the familiar von Neumann and Morgenstern utilities of conventional game theory, are called expected *rewards* to distinguish them from the *fitnesses* that will appear soon. The randomness that motivates the label "expected" will be introduced shortly.

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

Figure 1: The Reward Game $\mathcal{R}$

We assume that there is a single population containing N agents. Time is divided into discrete intervals of length τ. In each time period, an agent is characterized by the strategy X or Y that he is programmed to use in that period. In each period of length τ, pairs of agents are randomly (independently and with replacement) drawn 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.[3] An agent playing X, for example, will receive an

[3]Since we shall consider the case when $\tau \to 0$, 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. Similar hypotheses about rapid play are common in the literature, although they are sometimes hidden in the assumption that learning agents always switch to the current best reply. In particular, an agent cannot know the current best reply unless she can observe what the other agents are currently playing. Kandori, Mailath and Rob [**?**] are explicit in assuming that agents play an infinite number of times in each period. Nöldeke and Samuelson [**?**] assume a round-robin tournament in each period. Young's model [**?**] is less demanding in this respect, though agents still have access to the result of each game as soon as it is played.

expected reward of A in a population in which all agents play X and an "average" expected reward of $qA + (1-q)C$ when proportion q of his opponents play X and proportion $(1-q)$ play Y. In some cases, noted below, the model is formally identical to one in which each agent plays only once in each period.

Noisy learning. We consider a learning model that couples an aspiration-based rule for abandoning existing strategies with an imitation process for choosing new ones.[4]

At the end of each period of length τ, each agent experiences a Bernoulli trial which rings a mental bell with probability $\beta\tau$. (Without loss of generality, we subsequently take $\beta = 1$.) If the bell does not ring, the agent does not consider changing her strategy. If the bell rings, an event that we refer to as "receiving the learn draw", then the agent recalls her average realized reward in the last period and compares this with a fixed *aspiration level* Δ. The average realized reward in each period is random, being given by the sum of the average expected reward, denoted ρ and calculated from the Reward Game, and the realization R of an idiosyncratic random variable $\tilde{R}$ with cumulative distribution F.[5] If the average realized reward exceeds the aspiration level ($\rho + R > \Delta$), then the agent makes no change in strategy. If instead the average realized reward falls short of the aspiration level ($\rho + R < \Delta$), then the agent loses faith in her current strategy and abandons that strategy. We assume that F is log-concave, i.e., that $\ln F(z)$ is concave in z.[6]

If agent i has abandoned her strategy, then she must now choose a new strategy. We assume that she randomly selects a member j of the population. With

[4] Aspiration-based learning rules are examined by Bendor, Mookherjee and Ray [**?**] and Gilboa and Schmeidler [**?**, **?**, **?**].

[5] The somewhat awkward phrases "average expected reward" and "average realized reward" arise because the payoffs in the Reward Game are expected payoffs, to which R must be added to get realized payoffs; and because the relevant expected payoff is the *average* of the expected payoffs received from games played in the repeated matches with the various members of the population. The random variable $\tilde{R}$ represents an independent, identically distributed individual factor that that is independent of the current state and strategy and that yields a shock common to *each* payoff received by that agent in the given period. It would be interesting, but more complicated, to allow the distribution F to depend either 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. Papers in which this type of uncertainty appears include Fudenberg and Harris [**?**] and Robson [**?**].

[6] See Bagnoli and Bergstrom [**?**] for a discussion of log-concavity and its implications. We actually need log-concavity only on an open set containing $\{\Delta - \rho : \rho \in \{A, B, C, D\}\}$, and some of our examples will satisfy only this weaker assumption.

probability $1-2\lambda$, i imitates j's strategy.[7] With probability 2λ, i is a "mutant" who either does not observe j's strategy or ignores this observation, instead simply choosing a strategy randomly, with equal probability of choosing each strategy. We introduce mutations at this point in

the story, rather than in numerous other alternative places, because we believe mistakes are most likely to occur when one agent copies the strategy of another. Overall, i then imitates j with probability $1-\lambda$ and adopts instead the strategy that j is *not* playing with probability λ. We hereafter find it convenient to refer to i as

a mutant only in the case in which i adopts the strategy that j is not playing, and hence

to refer to λ as the mutation rate.

Aspiration Levels 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. 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.[8]

Fitness Game $\mathcal{F}$. The rewards that the players receive are not necessarily directly relevant to the dynamics that control the evolution of a population. Instead, it is important to translate these rewards ρ into *fitnesses* π. In general, how rewards are translated into fitnesses will depend on the dynamic process by which evolution proceeds.

In our model, the relevant fitnesses are the probabilities that a player who has received the learn draw will lose faith in and abandon his current strategy. We follow [**?**] by referring to these as *death probabilities*. For each expected reward ρ, the corresponding death probability is given by $\pi(\rho) = \text{prob}(\rho + R < \Delta) = F(\Delta - \rho)$; we say that the function $\pi : \mathbb{R} \to [0,1]$ is the death probability *induced* by the

[7] 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.

[8] 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.

distribution F. We refer to the process by which rewards are transformed into fitnesses (death probabilities) as the *fitness process*.

An interesting special case is that of a *uniform threshold rule with background fitness*. In this case, the random payoff shock $\tilde{R}$ takes on a value less than $\Delta - \overline{\rho}$, where $\overline{\rho} = \max_{\rho \in \{A,B,C,D\}}$, with probability θ and with probability $1 - \theta$ is drawn from a uniform distribution on an interval $[\,-\omega, \omega\,]$, where $\{A, B, C, D\} \subset [\Delta - \omega, \Delta + \omega]$. The parameter θ is the background fitness probability, with which a learning agent abandons a current strategy regardless of payoffs, and upon which we expand in Section 6. Death probabilities are linear in expected rewards in this case. We can then represent them in terms of a *Fitness Game*. In particular, let $\alpha = \text{prob}(A + R < \Delta) = F(\Delta - A)$, with similar

definitions for β, γ, and δ. Then the Fitness Game is given by Figure 2. The

	X	Y
X	α α	β γ
Y	γ β	δ δ

Figure 2: The Fitness Game

probability that strategy X is abandoned when x of $N - 1$ opponents play strategy X is given by $\alpha x/(N - 1) + \gamma(N - x)/(N - 1)$, which is the probability that the reward from strategy X falls below the aspiration level Δ. Notice that the Reward and Fitness Games have the same best-reply correspondence (recognizing that death is a "bad" rather than a "good", so that one wants to minimize payoffs in the Fitness Game). They therefore have the same payoff-dominant and risk-dominant equilibria. In this case, the model is equivalent to one in which each agent plays only once in each period.

For other distributions of $\tilde{R}$, death probabilities will be nonlinear in expected rewards, so we cannot represent them with a simple 2×2 fitness game. Nevertheless, a notion of risk dominance is still useful in such cases. We say that strategy Y is *risk dominant* in the fitness process given distribution F if:

$$\int_0^1 \pi(kB + (1-k)D)dk < \int_0^1 \pi(kA + (1-k)C)dk.$$

Hence, strategy Y is risk dominant if the expected probability of death for strategy X exceeds that of Y.[9] If the fitness process is represented by the fitness game, then this is the usual notion of risk dominance. We will say that the distribution F *preserves dominance* if the same strategy is risk dominant in the Reward Game and in the fitness process. Dominance is clearly preserved if the fitness process can be represented by a fitness game.

Limits. In Binmore, Samuelson and Vaughan [**?**] we argue that much of the work in constructing an evolutionary model occurs when one chooses the order in which to take the various limits involved in the model. Let t be the time at which the system is to be studied. Then we shall take four limits in this model, in the following order: $t \to \infty$, $\tau \to 0$, $N \to \infty$, and $\lambda \to 0$.

In letting $t \to \infty$ first, we are examining the stationary distribution reached by the system in the *ultralong* run. In particular, letting t go to infinity first ensures that we examine the stationary distribution without allowing other limiting operations, most notably that of $N \to \infty$, to obscure the random shocks that drive

the ultralong-run selection results.[10] In letting $\tau \to 0$, we are approximating the case of continuous time. We do this because we think it appropriate to study the case in which births occur at random intervals, as in an overlapping generations model. This leads us to concentrate on the case when τ is small compared with both N and λ, or $\tau \to 0$.

In letting $N \to \infty$ and, $\lambda \to 0$, we plan to study the case of large populations with small mutation rates. We do not see obvious principles guiding the choice of the order in which to take these limits, provided that they are taken after $t \to \infty$ and $\tau \to 0$. Fortunately it sometimes does not matter to the equilibrium selected, as in the case of a symmetric 2×2 game with two strict Nash equilibria. When it does matter, we allow $N \to \infty$ before $\lambda \to 0$. The reverse order yields a system that can get "stuck" in a state where one strategy has accidentally become extinct

[9] The expectation is taken over all (equally weighted) proportions of the population that might be playing strategy X, so that the expected probability of death for strategy X is $\int_0^1 (\pi(kA+(1-k)C)dk$.

[10] Binmore, Samuelson and Vaughan [**?**] show that if we first let τ approach zero and then N approach infinity, we obtain a continuous time, deterministic version of the replicator dynamics that provides a useful approximation of the *long-run* behavior of our muddling model. The replicator dynamics but do not provide an adequate model of the *ultralong* run. Instead, by taking the limit $N \to \infty$ before $t \to \infty$, the replicator dynamics "smooth out" random perturbations that are very unlikely to occur in the long run, and hence are not important to a long run analysis, but are crucial to ultralong-run considerations.

and cannot reappear.

3 Dynamics

Stationary Distribution. For a fixed set of values of the parameters τ, λ, and N, we have a homogeneous Markov process on a finite state space. Let $x \in \{0, 1, \ldots N\}$ specify the current state of the system, where x denotes the number of agents who are currently playing strategy X. Given any state $x \in \{0, 1, \ldots N\}$, 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 transition matrix for the Markov process is thus irreducible. The following result is then standard:

Proposition 1 *The Markov process has a unique stationary distribution. The expected frequencies of the states along any realization of the Markov process converge to this distribution; and the distribution of states at a given time t converges to this distribution, as t approaches infinity, from any initial condition.*

Proof. Kemeny and Snell [**?**], Theorems 4.1.4, 4.1.6, and 4.2.1. □

We study the stationary distribution in the limit as $\tau \to 0$. Binmore, Samuelson and Vaughan [**?**] consider the details of this limiting analysis and provide the details of the proof of Proposition 2 below. The relevant implication for our analysis is that we can work with arbitrarily short time periods, in the sense that in a single period, there will be either no agents who receive the learn draw or one agent who receives the learn draw. The event that more than one agent receives the learn draw occurs with negligible probability and can be ignored. Hence, given that the current state is x, the only transitions with which we need be concerned are to states x, $x + 1$, and $x - 1$.

Given state x, we let $r(x)$ be the probability that the system moves to state $x + 1$, i.e., moves one state to the right. Similarly, $\ell(x)$ is the probability of moving to state $x - 1$, or one state to the left. Consider $r(x)$. For the number of agents playing X to increase, four events must occur: (1) An agent must receive the learn draw. Let Λ be the probability with which this occurs. (2) The agent who receives the learn draw must be playing strategy Y, since the number of agents playing X can increase only if one agent abandons 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 $(N - x)/N$. (3) The learning agent must abandon his current strategy. Because the average payoff of an agent playing strategy Y is $(xB + (N - x - 1)D)/(N - 1)$, this occurs with probability $\pi((xB + (N - x - 1)D)/(N - 1))$. (4) The learning agent must choose X for his 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:

$$r(x) = \Lambda \frac{N - x}{N} \pi \left(\frac{xB + (N - x - 1)D}{N - 1} \right) \frac{(1 - \lambda)x + \lambda(N - x - 1)}{N - 1}. \qquad (1)$$

The value of $\ell(x)$ is defined analogously.

The basic tool used to characterize the stationary distribution is the following:[11]

Proposition 2 *Consider states x and $x + 1$. Let $\sigma(x)$ be the probability attached to state x by the stationary distribution of the Markov process. Then:*

$$\frac{\sigma(x + 1)}{\sigma(x)} = \frac{r(x)}{\ell(x + 1)}. \qquad (2)$$

Proof Sketch. From Freidlin and Wentzell ([**?**], Lemma 3.1 on page 177), $\sigma(x + 1)/\sigma(x)$ is given by the ratio of the sum of the products of the transition probabilities attached to all "$x + 1$-trees" to the similar calculation for "x-trees". Because only a single agent can receive the learn draw in each period, however,

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).[12] It is then apparent that these two trees differ only in one probability: The $x+1$-tree contains the probability $r(x)$ while the x-tree contains $\ell(x+1)$, giving (2). □

[11] From (1), we see that the probability Λ does not appear in (2), so that we need not discuss the details of the determination of

Λ here.

[12] The complete argument here would retain the very small probabilities of multiple learn draws in each period, but would then observe that 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.

To interpret this result, notice that if λ is very small and N very large (we will see that the order of limits is interchangeable here), (2) gives:

$$\frac{\sigma(x+1)}{\sigma(x)} \approx \frac{\pi_Y(x/N)}{\pi_X(x/N)}, \tag{3}$$

where $\pi_X : [0,1] \to \mathbb{R}$ and $\pi_Y : [0,1] \to \mathbb{R}$ are the death probabilities of strategies X and Y, in the limit as N gets large, given that proportion x/N of the population is playing strategy X.[13] Now consider a game with two strict Nash equilibria. Let x^*/N be the probability attached to X by the mixed-strategy, Nash equilibrium of the game. (Note that x^* need not be an integer.) Then $\sigma(x+1) > \sigma(x)$ whenever $x > x^*$ (because strategy X must be a best reply here, if the game is to have two strict Nash equilibria, and hence the ratio of death probabilities satisfies $\pi_Y/\pi_X > 1$). The stationary distribution σ must then increase on $[x^*, N]$. Similarly, $\sigma(x+1) < \sigma(x)$, and $\sigma(x)$ must decrease, on $[0, x^*]$. The graph of σ then has an "inverted bowl" shape, reaching its highest points at the endpoints of the state space, where we have the strict Nash equilibria in which

either all agents play X or all agents play Y, and its minimum at x^*.

Convergence. How long is the ultralong run? We first provide a comparison of the convergence properties, for small mutation probabilities, of our muddling model and the model of Kandori, Mailath and Rob [**?**]. We consider the case of a game with two strict Nash equilibria. We fix the value of N and examine the performance of the system for very small values of λ.

Let $\sigma^*_{KMR}(\lambda)$ be the stationary distribution of the Kandori, Mailath and Rob model given mutation rate λ. Consider the following measure, which is examined by Ellison [**?**]:

$$\sup_{\sigma_0} \limsup_{t \to \infty} \| \sigma_0[\Gamma_{KMR}(\lambda)]^t - \sigma^*_{KMR}(\lambda) \|^{\frac{1}{t}},$$

where σ_0 is the initial distribution and $\Gamma_{KMR}(\lambda)$ is the transition matrix of the Kandori, Mailath and Rob Markov process given mutation rate λ. This is then a measure of the distance between the distribution at time t (given by $\sigma_0[\Gamma_{KMR}(\lambda)]^t$) and the stationary distribution given λ (given by $\sigma^*_{KMR}(\lambda)$). Ellison shows that

[13] Hence, $\pi_X(x/N) = \pi((x/N)A + ((N-x)/N)C)$, $\pi_Y(x/N) = \pi((x/N)B + ((N-x)/N)D)$. If it seems paradoxical that it is the average payoff to strategy Y that appears in the numerator of (3), recall that these payoffs represent death probabilities and are hence "bads" rather than "goods".

there exists a function $\Theta_{KMR}(\lambda^z): \mathbb{R} \to \mathbb{R}$ such that

$$1 - \sup_{\sigma_0} \limsup_{t \to \infty} \| \sigma_0[\Gamma_{KMR}(\lambda)]^t - \sigma^*_{KMR}(\lambda) \|^{\frac{1}{t}} = \Theta_{KMR}(\lambda^z) \sim \lambda^z, \tag{4}$$

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. (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, $c|g(\lambda)| \leq |f(\lambda)| \leq C|g(\lambda)|$). The result (4) then gives us a measure of how fast the Kandori, Mailath and Rob model converges.

We can provide some intuition for why a term of order λ^z appears when considering convergence in the Kandori, Mailath and Rob model. Let all agents initially play the strategy that has the smaller basin of attraction.[14] Then the probability of switching from the state with the small basin of attraction to the state with the large basin of attraction in a single period is given by the probability of at least z mutations, or:

$$\sum_{h=z}^{N} \binom{N}{h} \lambda^h (1-\lambda)^{N-h}. \tag{5}$$

A lower bound on the speed of convergence of the Kandori, Mailath and Rob model is then given by the fact that it cannot be very close to its stationary distribution until sufficient time has elapsed for an event to have occurred whose probability in any given period is on the order of λ^z.

We now seek an analogous measure of the rate at which our muddling model converges. First, 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 basic 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 discrete Kandori, Mailath and Rob model and our continuous model comparable, we choose the units in which time is measured so that in the Kandori,

[14] It may seem misleading to take such an extreme initial condition. Notice, however, that the best–response learning scheme immediately moves the system to this state from any initial condition in the basin of attraction of this equilibrium, so that this initial condition is relevant for a potentially large proportion of the state space.

Mailath and Rob model every agent learns at each of the discrete times $\{1, 2, \ldots, \}$. We then let β, the probability of a birth per unit time in our model, equal one, so that, in the limit as our time period length τ shrinks, the expected number of times in an interval of time of length one (which will contain very many of our very short time periods) that an agent in our model learns is one, matching the Kandori, Mailath and Rob model.

Let $\Gamma_M(\lambda, \tau)$ be the transition matrix for the Markov process of our muddling model given mutation rate λ and period length τ. The quantity $\Gamma_M(\lambda, \tau)$ depends on τ because the probability of an agent receiving the learn draw depends on the period length. Notice also that as τ decreases, the number t/τ of periods that occur by time t increases.

Proposition 3 *There exists a function Θ_M such that*

$$\lim_{\tau \to 0} \left(1 - \sup_{\sigma_0} \limsup_{t \to \infty} \| \sigma_0 [\Gamma_M(\lambda, \tau)]^{\frac{t}{\tau}} - \sigma_M^*(\lambda) \|^{\frac{1}{t-1}} \right) \leq \Theta_M(\lambda) \sim \lambda. \qquad (6)$$

The proof is contained in the Appendix. Together, (4) and (6) 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 $T_{KMR}(\eta)$ be the length of time required for the Kandori, Mailath and Rob model to be within η of its stationary distribution. Let $T_M(\eta)$ be similarly defined for our muddling model. Then from (4) and (6), we have $\eta = (1 - \Theta_{KMR}(\lambda^z))^{T_{KMR}(\eta)}$ and $\eta \geq (1 - \Theta_M(\lambda))^{T_M(\eta)-1}$, giving, for small values of λ,

$$\frac{T_{KMR}(\eta)}{T_M(\eta) - 1} \geq \frac{ln(1 - \Theta_{KMR}(\lambda^z))}{ln(1 - \Theta_M(\lambda))} \approx \frac{\Theta_M(\lambda)}{\Theta_{KMR}(\lambda^z)} \sim \frac{1}{\lambda^{z-1}}. \qquad (7)$$

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 [**?**] 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.

The difference in rates of convergence for these two models will be most striking when the mutation rate is very small. There remains the question of how the models might compare for less extreme rates of mutation. In [**?**], 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 3×10^{5}.

This result appears because the Kandori, Mailath and Rob model relies upon mutations to accomplish its transitions between equilibria, so that convergence requires waiting until the number of simultaneous mutations required to move from one equilibrium to the basin of attraction of another becomes a reasonably likely event. In contrast, the muddling model requires mutations only to escape boundary states. Once a single mutation has allowed this escape, then subsequent adjustments can be performed by the noisy learning dynamics. When mutation rates are small, the learning dynamics proceed at a very much faster rate than mutations occur, so that convergence requires waiting only long enough to make a single mutation

a reasonably likely event.

We can provide an additional perspective by noting that an analogous argument establishes:

Proposition 4 *For sufficiently small* λ*,*

$$\lim_{N \to \infty} T_{KMR}(\eta)/T_M(\eta) = \infty.$$

We thus have faster convergence for either small mutation rates or large population sizes.

4 Equilibrium Selection

We now consider equilibrium selection. To do this, we examine the stationary distribution of the Markov process, in the limit as the population size gets large and the mutation rate small. In particular, we begin with a stationary distribution and then study the limits $N \to \infty$ and $\lambda \to 0$ as an exercise in comparative statics. The order in which these

two limits are taken is one of the issues to be examined.

Two Strict Nash Equilibria. We first assume $A > B$ and $D > C$, so that the Reward Game has two strict Nash equilibria . Let $\sigma_{(\lambda,N)}(x)$ denote the stationary distribution of the Markov process on $\{0, 1, \ldots, N\}$ given the mutation rate λ and population size N. Extend this to a measure on $([0,1], \Im)$ (where $\Im$ is the Borel sets) by letting $\sigma_{\lambda,N}(A) = \sum_{(x/N)\in A} \sigma_{\lambda,N}(x)$. Then:

Proposition 5 *There exists a unique probability measure σ^* on $([0,1], \Im)$ with $\lim_{N\to\infty} \lim_{\lambda\to 0} \sigma_{(\lambda,N)} = \lim_{\lambda\to 0} \lim_{N\to\infty} \sigma_{(\lambda,N)} = \sigma^*$, where the limits refer to the weak convergence of probability measures. In addition, $\sigma^*(\{0\}) + \sigma^*(\{1\}) = 1$.*

Proof. We first construct a candidate for σ^*. Fix N. Then it must be the case that $\lim_{\lambda\to 0} \sigma_{(\lambda,N)}(0) + \lim_{\lambda\to 0} \sigma_{(\lambda,N)}(1) = 1$, because[15]

$$\lim_{\lambda\to 0} \frac{r(0)}{\ell(1)} = \lim_{\lambda\to 0} \frac{\ell(N)}{r(N-1)} = 0,$$

and the result then follows from (2) and the fact that $\lim_{\lambda\to 0}(r(x)/\ell(x+1))$ is nonzero and finite for every value $x \in \{1, 2, \ldots, N-2\}$. This result states that as the mutation rate approaches zero, the system 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 $\sigma^*(0) + \sigma^*(1) = 1$, 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:

$$\frac{\sigma_{(\lambda,N)}(1)}{\sigma_{(\lambda,N)}(0)} = \prod_{x=0}^{N-1} \frac{r(x)}{\ell((x+1))} =$$

$$\prod_{x=0}^{N-1} \frac{(x(1-\lambda) + \lambda(N-x-1))\,(N-x)}{((N-x-1)(1-\lambda) + \lambda x)\,(x+1)} \;\Xi = \Xi,$$

where the last equality is obtained by pairing each term corresponding to x in the numerator on the left side with the term in the denominator corresponding to $N - x - 1$, and where

$$\Xi = \prod_{x=0}^{N-1} \pi\left(\frac{xB + (N-x-1)D}{N-1}\right)\left(\pi\left(\frac{xA + (N-x-1)C}{N-x}\right)\right)^{-1}. \qquad (8)$$

[15]To conserve on notation, we freely write "0" and "1" as arguments of measures, rather than "$\{0\}$" and "$\{1\}$".

We then take logarithms to obtain:

$$\ln \Xi = \sum_{x=0}^{N-1} \left\{ \ln \pi \left(\frac{xB + (N-x-1)D}{N-1} \right) - \ln \pi \left(\frac{(x-1)A + (N-x)C}{N-1} \right) \right\}.$$

The Lebesgue dominated convergence theorem then allows us to conclude that, for any value of λ:

$$\lim_{N\to\infty} \frac{1}{N} \ln \frac{\sigma_{(\lambda,N)}(1)}{\sigma_{(\lambda,N)}(0)} = \int_0^1 (\ln \pi_Y(k) - \ln \pi_X(k)) dk. \tag{9}$$

Our candidate for σ^* thus gives (in addition to $\sigma^*(0) + \sigma^*(1) = 1$) that $\sigma^*(0) = 1$ if the right side of (9)

is negative and $\sigma^*(1) = 1$ if the right side of (9) is positive.[16]

Next, we argue that $\lim_{N\to\infty} \lim_{\lambda\to 0} \sigma_{\lambda,N} = \sigma^*$, where the limits are understood to represent weak convergence. From Theorem 2.2 of Billingsley [**?**], it suffices to show this for any relatively open subinterval of $[0,1]$. Fix such an interval A. If $\{0,1\} \not\subseteq A$, then the result follows because $\lim_{\lambda\to 0} \sigma_{\lambda,N}(A) = 0 = \sigma^*(A)$. If either 0 or 1 is contained in A, then the result follows because (9) holds for all λ.

It remains to show that $\lim_{\lambda\to 0} \lim_{N\to\infty} \sigma_{(\lambda,N)} = \sigma^*$. Toward this end, we note that, because (9) holds for any λ,

$$\frac{\lim_{\lambda\to 0} \lim_{N\to\infty} \sigma_{(\lambda,N)}(1)}{\lim_{\lambda\to 0} \lim_{N\to\infty} \sigma_{(\lambda,N)}(0)} = \frac{\sigma^*(1)}{\sigma^*(0)}.$$

An analogous argument shows that this inequality holds when $\{1\}$ and $\{0\}$ are replaced by sets of the

form $(1-\theta, 1]$ and $[0, \theta)$ for any $\theta < 1/2$. It then remains only to show that $(\lim_{\lambda\to 0} \lim_{N\to\infty} \sigma_{(\lambda,N)}(\{0,1\}) = 1$. For any integer n, consider the sets $[0, 1/n]$ and $[m/n, (m+1)/n]$, for $m \geq 2$ and $(m+1)/n < x^*/N$ (recall that x^*/N is the probability attached to strategy X by the mixed strategy equilibrium). Then

$$\frac{\sigma_{(\lambda,N)}([0, \frac{1}{n}])}{\sigma_{(\lambda,N)}([\frac{m}{n}, \frac{m+1}{n}])} \geq \prod_{x=N/n}^{Nm/n} \frac{r(x)}{\ell(x+1)}.$$

For sufficiently small λ, every term in the product on the right side of this inequality is less than one. This ensures that, for sufficiently small λ, $\lim_{N\to\infty} \sigma_{(\lambda,N)}([m/n, (m+$

[16] If the right side of (9) equals zero, then both $\sigma^*(0)$ and $\sigma^*(1)$ may be positive.

$1)/n]) = 0$. A similar argument applies to subsets of $[x^*/N, 1]$ and yields the result. □

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 the right side of (9) is nonzero), probability will be attached to only one of these equilibria, which we refer to as the "selected" equilibrium. From (9), we immediately have a criterion for which equilibrium will be selected:

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

$$\int_0^1 \ln \pi_Y(k)dk \quad > \quad [<] \quad \int_0^1 \ln \pi_X(k))dk. \tag{10}$$

A number of papers have recently addressed the problem of equilibrium selection in symmetric 2×2 games. A common finding is that the risk-dominant equilibrium in the Reward Game is selected. Our process does not always select the risk-dominant equilibrium of either the Reward Game or the fitness process:

Corollary 1 (1.1) *If the payoff-dominant equilibrium in the Reward Game is also risk dominant (in the Reward Game), then the payoff-dominant equilibrium is selected.*

(1.2) *The payoff-dominant equilibrium in the Reward Game can be selected even if it fails to be risk dominant in both the Reward and Fitness Games.*

Proof. (1.1) Let (X, X) and (Y, Y) be risk-equivalent in the Reward Game, so that $A + C = B + D$, and let $A = D$. Then

(10) 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 the risk-equivalence of the two strategies). This leaves $\int_0^1 \rho_X(k)dk$ unchanged, where ρ_X is the expected reward to strategy X given that the proportion of agents playing X in the current state is k. The log-concavity of F then ensures that $\int_0^1 \ln(\pi_X(k))dk$ decreases. The left side of (10) then exceeds the right side and the payoff-dominant equilibrium (X, X) is selected. Next, note that adding a constant to A and C or subtracting a constant from D and B so as to also make (X, X) risk dominant increases the function $\ln \pi_Y(k) - \ln \pi_X(k)$ on $[0, 1]$, which can only strengthen the inequality in (10), and hence preserves the result that the payoff-dominant equilibrium is selected.

(1.2) Let F be a uniform distribution. Then we can represent the fitness process with a Fitness Game, an example of which is given

in Figure 3. Neither of the two pure strategy Nash equilibria, given by (X, X)

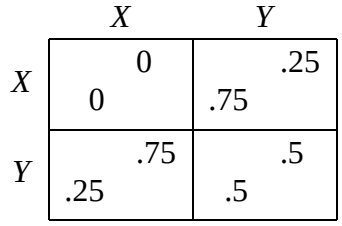

Figure 3: Fitness Game

and (Y, Y), risk-dominates the other in either the reward or fitness games, but (X, X) is the payoff-dominant equilibrium. The proof of (1.1) then indicates that (X, X) will be

the selected equilibrium. We can confirm this by performing the integration in (9) to give:

$$\frac{\sigma^*(1)}{\sigma^*(0)} = \lim_{N\to\infty} \left(\left(\frac{\delta^\delta}{\beta^\beta} \right)^{\frac{1}{\delta-\beta}} \left(\frac{\alpha^\alpha}{\gamma^\gamma} \right)^{\frac{1}{\gamma-\alpha}} \right)^N = \lim_{N\to\infty} \left(\frac{4}{3} \right)^N. \tag{11}$$

This game can now be perturbed so that Y is the risk-dominant equilibrium (and so that no death probabilities are zero); but with our model still selecting (X, X) and the latter continuing to be payoff-dominant but not risk-dominant in either the Reward or Fitness Games. □

We can provide some intuition as to why this result differs from that of Kandori, Mailath and Rob [**?**]. The Kandori, Mailath and Rob 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 given in Figure 3 that we considered in the previous proof, the equilibrium (X, X) has a basin of attraction which is smaller than (Y, Y)'s, but in (X, X)'s basin the relative payoff of X to Y is very small, being nearly zero for states in which nearly all agents play X. Given that payoffs are (undesirable) death probabilities, this makes it very difficult to leave (X, X)'s basin, and yields a selection in which all agents play X. Similar considerations appear in Fudenberg and Harris [**?**], who point out that the system

of Kandori, Mailath, and Rob moves away from an equilibrium by "jumping over" the relevant basin of attraction, so that only the size of this basin matters; whereas in a continuous system such as ours, the process must "swim upstream" through the basin of attraction, causing the "slope" as well as size of the basin to matter.

No Pure Strategy Equilibria. We can contrast these results with the case of games in which $B > A$ and $C > D$, so that there is a unique, mixed-strategy Nash equilibrium. Then an argument analogous to that of the previous two propositions gives:

Proposition 7 *Let x^*/N be the probability attached to X in the mixed strategy equilibrium. Then $\lim_{\lambda \to 0} \lim_{N \to \infty} \sigma_{(\lambda,N)}(A) = 0$ if $x^*/N \notin A$. However, $\lim_{\lambda \to 0} \sigma_{(\lambda,N)}(0) + \lim_{\lambda \to 0} \sigma_{(\lambda,N)}(1) = 1$.*

The order of limits makes a difference in this case. If mutation rates are first allowed to approach zero, then the long-run dynamics are driven by the possibility of accidental extinction coupled with the impossibility of recovery, attaching probability only to the nonequilibrium states in which all agents play the same strategy. 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 embrace the latter as the more useful model.

5 Comparative Statics

In this section, we establish several comparative static results for the case of games with two strict Nash equilibria.

High Background Fitness. Suppose that the random payoff variable $\tilde{R}$ is given by the value $\underline{R}$ with probability θ, where $\underline{R} + \overline{\rho} < \Delta$. (Recall $\overline{\rho} = max_{\rho \in \{A,B,C,D\}} \rho$). With probability $1 - \theta$, $\tilde{R}$ is given by a draw from a distribution $\hat{F}$. Hence, with probability θ, a learning agent simply abandons the current strategy regardless of its payoff. We say that the distribution of $\tilde{R}$ (denoted F) has *background fitness level* θ when it takes this form. In this section, we hold $\hat{F}$ fixed and investigate the implications of varying levels of background fitness θ.

The probability θ is intended to capture background considerations that are independent of what happens in the game under study. Such notions of background

fitness have played an important role in biological models of evolution. If $\theta = 1$, then the process is driven *entirely* by background considerations, in that decisions to switch strategies have nothing to do with the rewards in the game $\mathcal{R}$. If $\theta = 0$, then there are no background considerations and only rewards in the game matter when adjusting strategies.[17]

Proposition 8 *Let F have background fitness θ. Then for sufficiently large θ, the risk-dominant equilibrium of the fitness process is selected. If F preserves dominance, then for sufficiently large θ the risk-dominant equilibrium of the Reward Game is therefore selected.*

Our model thus justifies selecting the risk-dominant equilibrium in economic games as long as one believes that the learning context in which the game is played is such that strategy choices have little to do with the game.[18]

Proof. Let (X, X) and (Y, Y) be Nash equilibria, with (Y, Y) being risk dominant in the fitness process. From (10), the selected equilibrium will be the risk-dominant equilibrium if:

$$\int_0^1 (\ln(\theta + (1-\theta)\pi_Y(k)) - \ln(\theta + (1-\theta)\pi_X(k))dk < 0, \tag{12}$$

where π_Y and π_X are the death probabilities induced by $\hat{F}$. As

$\theta \to 1$, the left side of (12) approaches zero. We accordingly examine the first derivative of the left side of (12). A positive derivative ensures that (12) approaches zero from below, so that (12) is negative and hence the risk-dominant equilibrium is selected for θ near 1. Evaluated at $\theta = 1$, this derivative is:

$$\int_0^1 (\pi_X(k) - \pi_Y(k))dk. \tag{13}$$

[17]There are several alternative interpretations of the level of background fitness. For example, agents may be occasionally withdrawn from the game-playing population. Alternatively, bounded rationality might be at work, there being some probability that an agent's thought processes are currently congested with other considerations. In either case, it seems reasonable to assume that an agent's switching decision might then be determined by factors extraneous to the game. For 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 [**?**]. Strategy revisions will then often be determined by what happens in games other than the game actually under study.

[18]Notice that whether strategy Y is risk-dominant in the fitness process and hence whether dominance is preserved depends only on $\hat{F}$ and not on the level of background fitness, so that the proposition is well-defined in invoking risk-dominance notions independently of θ.

Because (Y, Y) is risk dominant in the fitness process, (13) is positive. The selected equilibrium, for large background fitness, is thus the equilibrium that is risk dominant in the fitness process. If F preserves dominance, then this is also the equilibrium that is risk dominant in the Reward Game. □

To see the forces behind this result, consider the function $\pi_Y(k)/\pi_X(k) \equiv \Psi(k)$, where π_Y and π_X are induced by F. From (3) and (9), this function holds the key to determining which equilibrium is selected. Recall that $\Psi(k)$ is increasing in k, and that a bias in favor of the payoff-dominant equilibrium is caused by the fact that values of $\Psi(k)$ are especially large for high values of x. The reason is that (X, X) is the payoff-dominant equilibrium, so the death probability π_X is small (and so Ψ large) for high values of x. As θ approaches unity, however, Ψ also converges to unity, because deaths are now caused primarily by exogenous factors rather than the payoffs of the Reward Game. The effect of having very large values of Ψ for values of x near one is then removed, and equilibrium selection is driven by the sizes of the basins of attraction. This leads to selection of the risk-dominant equilibrium.

In summary, we get the risk-dominant equilibrium in relatively unimportant games – those whose payoffs have little to do with agents' behavior. Notice that the relevant risk dominance notion arises in the fitness process and not in the Reward Game. This is simply a reflection of the fact that the learning process is driven by fitnesses rather than rewards, and the common practice of treating rewards as fitnesses may not always be appropriate.

Noisy Learning. Another result can be interpreted as saying that risk-dominant equilibria are selected in unimportant games. Say that the distribution F *has noise level* θ if F attaches probability $\theta/2$ to each of the values $\underline{R}$ and $\overline{R}$, with $\underline{R} + \overline{\rho} < \Delta$ and $\overline{R} + \underline{\rho} > \Delta$ (where $\underline{\rho} = \min_{\rho \in \{A,B,C,D\}} \rho$) and if F is given by the distribution $\hat{F}$ with probability $1 - \theta$. Hence, with probability θ, the agent simply chooses randomly whether to abandon or retain a strategy, with equal probability of each outcome. Then the proof of the previous proposition can be recycled to show:

Proposition 9 *Let F have noise level θ. Then for sufficiently large θ, the risk-dominant equilibrium of the fitness process is selected. If F preserves dominance, then for sufficiently large θ the risk-dominant equilibrium of the Reward Game is therefore selected.*

We again have the risk-dominant outcome appearing in games where agents' strategy choices have little to do with the payoffs of the game, and hence have risk dominance in relatively unimportant games.[19]

Best-Response Dynamics. We next ask how the selected equilibrium varies as the learning rule approaches best–response learning. Let $A > B$ and $D > C$, and let $x^*/N > 1/2$; so that there are two strict Nash equilibria, with (Y, Y) being the risk-dominant equilibrium. Let Δ equal the mixed-strategy equilibrium payoff, and let the distribution of $\tilde{R}$ put probability mass ϕ on zero and be otherwise distributed according to some function $\hat{F}$ with probability $1-\phi$. The death probability for strategy Y at state x is then given by $\phi + (1-\phi)\pi((xB+(N-x-1)D)/(N-1))$ if X is a best reply and by $(1-\phi)\pi((xB+(N-x-1)D)/(N-1))$ otherwise, where π is induced by $\hat{F}$. As ϕ approaches unity, the learning rule approaches the best-reply dynamics.

Proposition 10 *Let Δ equal the mixed strategy equilibrium payoff. Fix a distribution $\hat{F}$ and let F put a probability mass of $\phi \in (0,1)$ on zero and otherwise be given by $\hat{F}$. Then for values of ϕ sufficiently close to one, the selected equilibrium is the risk-dominant equilibrium in the Reward Game.*

Proof. Let $k^* \equiv x^*/N > 1/2$, so that (Y, Y) is the risk-dominant equilibrium in the Reward Game. From (10), the selected equilibrium will be (Y, Y) if:

$$\int_0^{k'} \{\ln((1-\phi)\pi_Y(k)) - \ln(\phi + (1-\phi)\pi_X(k))\}dk$$

$$+ \int_{k'}^{k^*} \{\ln((1-\phi)\pi_Y(k)) - \ln(\phi + (1-\phi)\pi_X(k))\}dk$$

$$+ \int_{k^*}^{1} \{\ln(\phi + (1-\phi)\pi_Y(k)) - \ln((1-\phi)\pi_X(k))\}dk < 0,$$

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. □

[19] Many variations on these last two propositions, involving various rules under which $\tilde{R}$ is chosen from a distribution $\hat{F}$ with probability $1-\theta$ and with probability θ is determined by a random process exogenous to the game, can be similarly established.

The forces that drive this result mimic those that appear in the models of Young [**?**] and Kandori, Mailath and Rob [**?**]. If ϕ is very close to 1, then the learning process almost certainly pushes the system in the direction of a best reply, and moving against this direction requires a mutation. As the mutation probability gets arbitrarily small, the only relevant factor in equilibrium selection becomes the number of mutations required to leave each equilibrium. This number is higher for the risk-dominant equilibrium in the Reward Game, and the latter is selected.

Risk Dominance. We now ask how the likelihood of choosing the risk-dominant equilibrium in the Reward Game varies with the rewards in that game. To pose this question precisely, let $k^* = x^*/N$ 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 Υ be the payoff in the Reward Game from the mixed strategy equilibrium. We will fix k^* and Υ and vary the rewards A,B,C,D. In particular, let $C(A)$ and $B(D)$ solve $(1-k^*)C(A) + k^*A = (1-k^*)B(D) + k^*D = \Upsilon$. We will then examine pairs of values (A,D), and take the associated B and C to be given by $B(D)$ and $C(A)$, which ensures that k^* and Υ (and hence the best-reply correspondence) are preserved as

rewards are varied. Let $\Omega(k^*,\Upsilon) = \{(A,D) : D \in (C(A),A)\}$; this is the set of possible values of A and D that are consistent with (X,X) being the payoff-dominant equilibrium. Let $\Omega^*(k^*,\Upsilon) \subset \Omega(k^*,\Upsilon)$ be the subset of values for which the payoff-dominant equilibrium in the Reward Game is selected. Then we have:

Proposition 11 (11.1) *Fix k^* and Υ. If $(A,D) \in \Omega^*(k^*,\Upsilon)$ and $(A',D') \in \Omega(k^*,\Upsilon)$ with $A' \geq A$ and $D' \leq D$, then $(A',D') \in \Omega^*(k^*,\Upsilon)$.*

(11.2) *Fix Υ. Then there exist values of k^* for which $\Omega^*(k^*,\Upsilon)$ is nonempty, and if $\Omega^*(k^{**},\Upsilon)$ is nonempty then $\Omega^*(k^*,\Upsilon)$ is nonempty for all $k^* \in (.5,k^{**})$.*

This proposition tells us, from (11.1), that movements in the rewards 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 is that in which $D < B$. Hence, the payoff-dominant equilibrium is favored by reducing the variation in rewards 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. On the other hand, the payoff-dominant equilibrium appears if A is large, so that there is considerable variation in the rewards to strategy X. Finally, from (11.2),

we have 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: (11.1) Fix $k^* > 1/2$. Notice that $C(A)$ is linear. The second derivative of $\int_0^1(\ln(\pi_X(k))dk = \int_0^1(\ln F(\Delta - ((1-k)C(A) + kA)))dk$ with respect to A is then given

by $\int_0^1(\zeta^2(F(\eta)F''(\eta) - (F'(\eta))^2)/(F(\eta))^2)dk$, where $\eta = \Delta - ((1-k)C(A) + kA)$ and $\zeta = d\eta/dk$. The concavity of $\ln F$ ensures that this is negative and hence $\int_0^1(\ln(\pi_X(k))dk$ (and similarly $\int_0^1(\ln(\pi_Y(k))dk$) is concave. Fix A and hence $C(A)$. It is clear that if we set $D = A$, then the equilibrium (Y, Y) will be selected, since it is risk dominant and payoff undominated. Now let D decline. From (10), the concavity of $\int_0^1(\ln(\pi_Y(k))dk$ ensures that if there are any values of D for which (X, X) is selected, they form an interval of the form $(C(A), D')$ for some $D' < A$. Similarly, fix D. If $A = D$, then (Y, Y) is selected. From (10), the convexity of $-\int_0^1(\ln(\pi_X(k))dk$ ensures that if there is a set of values for which the payoff-dominant equilibrium is selected, it is of the form (A', ∞) for some A'.

(11.2) 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 Υ. Let $1/2 < k^* < k^{**}$. Then we can find values $A' > A$, $C' > C$, $B' < B$, and $D' < D$ that (1) give mixed strategy equilibrium k^*; (2) increase all rewards to X and decrease all rewards to Y; and (3) preserve Υ. These payoffs must then preserve the property that the payoff-dominant equilibrium is selected, so $\Omega^*(k^*, \Upsilon)$ is nonempty. □

Given this result, it is interesting to note that Straub [**?**] has conducted experiments to investigate the conditions under which risk-dominant and payoff-dominant equilibria are selected. Straub found that the risk-dominant equilibrium was the most common equilibrium played in seven out of eight of the

experiments. The exception, in which the payoff dominant equilibrium appeared,

was the only game in which $D < B$.

6 Endogenous Learning

Our model is driven by a particular learning process that is arbitrary in some respects. Why should we be interested in this process, or more generally, why should we be interested in any evolutionary model that is driven by an arbitrary learning rule that is not derived from empirical studies?

One response to this question emphasizes the need to investigate the implications of many learning rules of different kinds. The analysis of the previous section is in this spirit. A potentially more fruitful approach, in our view, is to recognize that learning rules themselves are likely to have been learned in an evolutionary process. In this section, we therefore expand our model so that the learning rules themselves become the objects of evolutionary selection.

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, the results we establish below hold for any evolutionary process in which learning rules that yield higher average payoffs fare better than those with lower 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. If this robustness persists outside of our narrow model, then it may be unnecessary to proceed further.

In seeking to endogenize some aspects of the learning rule, we allow only the value of the aspiration level Δ to be subject to change.[20] We therefore label a learning rule by the aspiration level Δ which it incorporates.

Rational Expectations. One might expect the aspiration level Δ to bear some relationship to the

payoffs earned in the game. We accordingly first investigate a simple "rational expectations" property. We say that the learning process exhibits *rational expectations* if the aspiration level Δ is set equal to the average expected payoff of the selected equilibrium.

It may appear demanding for every agent to require at least an average payoff. However, there is no logical inconsistency in holding an aspiration level that not

[20]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.

only equals the average payoff but even equals a much higher value, though such an agent is likely to switch strategies frequently. Indeed, we are all familiar with colleagues who, driven by a quest for exceptionally high payoffs, change strategies (in this case, perhaps, jobs) frequently. The use of such an

aspiration level should not be confused with the inconsistent behavior of the vast majority of drivers who consider themselves to be better than average at driving, or the academic departments which demand that their members all receive above average teaching ratings. Finally, similar impossibility results could be constructed that require not that the aspiration level be equal to the average payoff but merely that it be positively related to the average payoff.

Appendix II proves:

Proposition 12 *There exist games in which it is impossible for the learning process to exhibit rational expectations.*

To see why this impossibility appears, let A be the payoff from the payoff-dominant equilibrium and D the payoff from the risk-dominant equilibrium, so $A > D$. Setting the aspiration level equal to A yields relatively high death probabilities for both strategies. However, relatively high death probabilities for both strategies are the conditions under which the risk-dominant equilibrium is likely to be selected. We have seen this in the results of Propositions 8 and 9, where high death probabilities for both strategies, produced, by high background fitness levels or noisy learning, drive the outcome to the risk-dominant equilibrium. Similarly, setting the aspiration level equal to D yields relatively low death probabilities for both strategies, which are the conditions under which the payoff-dominant equilibrium is likely to be selected. It is then possible that when $\Delta = D$, death probabilities are low enough to support the payoff-dominant equilibrium but setting $\Delta = A$ raises death probabilities enough to tip the system to the risk-dominant equilibrium. If so, it cannot be that Δ is set equal to the equilibrium payoff.

Evolution of Learning Rules. Proposition 12 prompts us to examine models in which the aspiration level Δ is set by evolutionary forces. Only games with two strict Nash equilibria are considered. Because we identify learning rules with their associated value of Δ, we refer to the evolution of Δ as the evolution of a learning rule.

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. We shall not offer an explicit dynamic model of the evolution of learning rules. Instead, we seek conditions under which a learning rule will satisfy conditions analogous to those of evolutionary stability.[21] These will ensure that our results hold for a broad class of evolutionary processes, namely those driven by differences in the average payoffs of various

learning rules.

We find that the evolution of the aspiration level Δ depends upon the distribution of $\tilde{R}$, and especially on how dispersed is this distribution. We first prove a simple result:

Proposition 13 *Consider an aspiration level Δ. Fix $t > 0$. Suppose that prob$(R < \Delta - t)$/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 this, consider the condition that prob$(R < \Delta - t)$/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 prob$(R > s + t)$/prob$(R > s)$ is decreasing in s. We can then calculate:

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

and it suffices to show:

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

To establish this inequality, we note the following:

$$\text{prob}(R > s + t) = \int_{s+t}^{\infty} e^{-x^2/2}dx = \int_{s}^{\infty} e^{-(t+x)^2/2}dx \; =$$

[21] See Harley [**?**], Maynard Smith [**?**], and Ellison and Fudenberg [**?**] for work in this vein.

$$e^{-t^2/2}\int_s^\infty e^{-tx}e^{-x^2/2}dx \leq e^{-t^2/2}e^{-st}\int_s^\infty e^{-x^2/2}dx = e^{-t^2/2}e^{-st}\text{prob}(R > s). \qquad (14)$$

Proof of Proposition 13. Let the agents in the population be distributed among 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 given by

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

where $\Pi(P_Y)$ is the expected payoff difference between aspiration levels Δ' and Δ conditional on the system being in set P_Y, and $\Pi(P_X)$ and $\Pi(P_D)$ are analogous. Because the time spend in set P_D can be made arbitrarily small by increasing N and decreasing λ, it suffices to show that at least one of $p_Y(N,\lambda)\Pi(P_Y)$ or $p_X(N,\lambda)\Pi(P_X)$ is positive and 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 $p_Y(N,\lambda)$ is, and consider $\Pi(P_Y)$. This is given by

$$\sum_{k=0}^{\infty}\left(\frac{kp_Y^k(N,\lambda)}{\sum_{h=0}^{\infty}hp_Y^h(N,\lambda)}\right)\Pi^k(P_Y), \qquad (15)$$

where $p_Y^k(N,\lambda)$ is the probability that a given episode in which the Markov chain 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:

$$\lim_{N\to\infty}\sum_{h=0}^{\infty}hp_Y^h(N,\lambda) = \infty, \qquad (16)$$

since, as N gets large, the system spends an arbitrarily small proportion of its periods 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 $\Pi^{\infty}(P_Y)$. Suppose $\Pi^{\infty}(P_Y) > 0$. Then for any $\epsilon > 0$, there is a $k' > 0$ such that for all $k > k'$,

$$\Pi^{k}(P_Y) > \Pi^{\infty}(P_Y) - \epsilon. \tag{17}$$

Then (17) and (16) (along with $\Pi^{\infty}(P_Y) > 0$) imply that (15) is positive for

sufficiently large N and small enough λ, and does not approach zero as N grows and λ shrinks, giving the result.

It then remains only to show that $\Pi^{\infty}(P_Y)$ is positive when $\Delta' < \Delta$. For this, however, it suffices that $\text{prob}(R < \Delta - t)/\text{prob}(R < \Delta)$ is increasing in Δ. In particular, within set P_Y, the notions of inferior and best reply are unambiguous; 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 of 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 $\text{prob}(R < \Delta - t)/\text{prob}(R < \Delta)$ 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. □

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 $\text{prob}(R < \Delta - t)/\text{prob}(R < \Delta)$ is increasing in Δ, these learning rules involve smaller aspiration levels.

If $\text{prob}(R < \Delta - t)/\text{prob}(R < \Delta)$ 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.[22] Whereas we have presented this result in terms of only two coexisting aspiration levels, it generalizes in a straightforward fashion to any existing configuration of aspiration levels. 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 (the extension to other Normal distributions is immediate).

Proposition 14 *Let F be the Standard Normal distribution. Then for sufficiently small Δ, the selected equilibrium is the equilibrium that is risk-dominant in the Reward Game.*

Proof: Let $A > C$ and $D > B$ with $A + C < B + D$, so that (Y, Y) is risk dominant. From (10)), we can calculate that in order to show that the risk-dominant equilibrium will be selected it suffices to show that $\ln F(\Delta - B) + \ln F(\Delta - D) - \ln F(\Delta - A) - \ln F(\Delta - C) < 0$. Let $N_A = 1 - F(-(\Delta - A))$, and let N_B, N_C, and n_D be analogous. 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'Hopital's rule can be used to show that $\lim_{s\to\infty} N_{s+t}/N_s = e^{-t^2/2}e^{-st}$. 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

[22] 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.

involving $-\Delta$, which gives $2(-\Delta)(A+C-B-D) < 0$. This will hold for large $-\Delta$ because (Y,Y) is risk dominant in the Reward Game, so that $A+C-B-D<0$ and hence the coefficient on $-\Delta$ is negative. □

Evolution thus leads to learning rules that will select the risk-dominant equilibrium in the Reward Game. Three considerations arise in interpreting this result.

First, there are widely dispersed distributions, such as the Cauchy distribution, which do not push Δ ever lower, and hence the payoff-dominant equilibrium may survive. To see that Δ is not relentlessly pushed lower under the Cauchy distribution, it it is sufficient to show that $\text{prob}(R > s+t)/\text{prob}(R > s)$ *increases* in s. Here, we have $\text{prob}(R > s) = \int_s^\infty (1/(\pi(1+x^2)))dx = (1/2) - (1/\pi)\arctan x$. We then need that

$$\frac{\pi/2 - \arctan(s+t)}{\pi/2 - \arctan s}$$

is increasing in s; and the argument is completed by noting that this ratio increases to unity as s increases. In this case, Δ will not be pushed ever-lower, since if Δ is sufficiently small, then further lowering Δ will increase $\text{prob}(R < \Delta - t)/\text{prob}(R < \Delta)$, lowering payoffs. At the same time, however, Δ will not be pushed arbitrarily high, since as Δ increases all death probabilities approach unity, eventually again increasing $\text{prob}(R < \Delta - t)/\text{prob}(R < \Delta)$ and lowering payoffs. There will then be a finite, payoff-maximizing value of Δ which may allow either the risk-dominant or

payoff-dominant equilibrium to survive.

Second, in cases where evolution produces a learning rule that in turn produces the risk-dominant equilibrium, we must consider the length of time required to produce such an outcome. We have described the evolution of learning rules as taking place at a slower pace than the evolution of play in the game. This suggests that we are talking about a time span even longer than the ultralong run, which is likely to be too long to be of relevance. Notice, however, that the arguments in the proof of Proposition 13 require only that the system spend most of its time close to either the risk-dominant or payoff-dominant equilibrium, and do not depend upon the relative amounts of time spent near these two equilibria. This

is a condition that is satisfied in the long run, allowing the evolution of learning rules to proceed in the ultralong run alongside the evolutionary process that produces our equilibrium selection results. There may then still be time for the evolution of learning rules to be relevant.

Finally, we have examined the evolution of a very narrow class of learning rules, allowing only the aspiration level to adjust. The task remains of examining

the implications of allowing evolution to affect other aspects of the model, such as the distribution of $\check{R}$, which might capture changes in the ability to identify the links between payoffs and actions. The task also remains of investigating the evolution of learning rules in fundamentally different learning models.

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 explore four issues that arise in the course of constructing such microfoundations.

First, it is important to identify the time-period of interest. The techniques

relevant for a long-run analysis are not the same as those required for an ultralong run investigation. In particular, perturbations of the learning process that are irrelevant in the long run assume center stage when analyzing ultralong-run selection results.

Second, the nature of these perturbations is important. Our innovation here is

to examine 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 third issue brought to the fore when constructing an evolutionary model is the difference between fitnesses and rewards. The payoffs that drive an evolutionary model are fitnesses. We think it important to bear in mind that the relationship between

fitnesses and the rewards that conventionally appear in games depends on the nature of the selection process lying behind the

evolutionary model.

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.

8 Appendix I: Proof of Proposition 3