A Numerical Analysis of the Evolutionary Stability of Learning Rules

Jens Josephson*
Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance No. 474
November 15, 2001

Abstract

In this paper I define an evolutionary stability criterion for learning rules. Using Monte Carlo simulations, I then apply this criterion to a class of learning rules that can be represented by Camerer and Ho's (1999) model of learning. This class contains perturbed versions of reinforcement and belief learning as special cases. A large population of individuals with learning rules in this class are repeatedly rematched for a finite number of periods and play one out of four symmetric two-player games. Belief learning is the only learning rule which is evolutionarily stable in almost all cases, whereas reinforcement learning is unstable in almost all cases. I also find that in certain games, the stability of intermediate learning rules hinges critically on a parameter of the model and the relative payoffs.

Keywords: Bounded rationality, Evolutionary game theory, Evolutionary Stability, Learning in games, Belief learning, Reinforcement learning.

JEL classification: C72, C73.

*This paper arose out of extensive discussions with Jörgen W. Weibull, who also provided many valuable comments on earlier drafts. I also want to thank Maria Sáez-Martí, Ariel Rubinstein and seminar participants at the Microeconomics Workshop of Cornell University, The Society for Economic Dynamics 2001 Annual Meeting, and the 16th Annual Congress of the European Economic Association, 2001. Part of this work has been completed at the Department of Economics of Cornell University, which I thank for its hospitality. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1 Introduction

The bounded rationality paradigm is based on the assumption that people learn to play games by using simple rules of adaptation, often referred to as *learning rules*. The objective is generally to predict which strategies are more likely to be observed in the long run, given that all players use a specific learning rule. A problem with this setting is that the learning rule is treated as exogenous and that no motivation is provided for the particular choice of learning rule. Evolutionary forces are usually only allowed on the level of simple strategies and not on the higher level of learning rules. In this paper, I attempt to take a step towards closing this open-endedness of the paradigm by developing an evolutionary stability criterion for learning rules and applying this criterion to a set of well-known learning rules using Monte Carlo simulations.

More specifically, I ask if there is a rule such that if applied by a homogeneous population of individuals, it cannot be invaded by mutants using a different rule. I call such an uninvadable rule an *evolutionarily stable learning rule* (ESLR). This concept is an extension of the classical definition of evolutionarily stable strategies (Maynard Smith and Price (1973), Maynard Smith (1974)) to learning rules and dynamic strategies.

The setting is a world where the members of a large population, consisting of an even number of individuals, in each of a finite number of periods are all randomly matched in pairs to play a finite two-player game.[1] Each individual uses a learning rule, which is a function of her private history of past play, and fitness is measured in terms of expected average payoff. This framework provides a rationale for the use of learning rules and it is of particular interest

[1] To check the robustness of my results, I have also analyzed a different matching scheme, where the individuals only are randomly matched at the start of the first period, and then continue to play against the same opponent for a finite number of periods. The results from the simulation of this matching scheme are in general consistent with the results for the the matching scheme in this paper.

since very little analysis of learning in this "repeated rematching" context has previously been done.

Technically, learning rules are mappings from the history of past play to the set of pure or mixed strategies. There are many models of learning and I therefore restrict the numerical analysis to a class of learning rules that can be described by the general parametric model of Camerer and Ho (1999), called *experience-weighted attraction learning* (EWA). The rules in this class have experimental support and perform well in an environment where the game changes from time to time. Moreover, the class contains rules which differ considerably in their use of information. Two of the most well-known learning rules, *reinforcement learning* and *fictitious play* (or *belief learning*), are special cases of this model for specific parameter values.

Reinforcement learning is an important model in the psychological literature on individual learning. It was introduced by Bush and Mosteller (1951) although the principle behind the model, that choices which have led to good outcomes in the past are more likely to be repeated in the future, is due to Thorndike (1898). Under reinforcement learning in games, players assign probability distributions to their available pure strategies. If a pure strategy is employed in a particular period, the probability of the same pure strategy being used in the subsequent period increases as a function of the realized payoff. The model has very low information and rationality requirements in the sense that individuals need not know the strategy realizations of the opponents or the payoffs of the game; all that is necessary is knowledge of player-specific past strategy and payoff realizations.

Fictitious play, or belief learning, is a model where the individuals in each of the roles of a game in every period play a pure strategy that is a best reply to the accumulated empirical distribution of their opponents' play. This means that knowledge of the opponents' strategy realizations and the player's own payoff function is required.

Several different models of both reinforcement and fictitious play have

been developed over the years. The ones that can be represented by Camerer and Ho's (1999) model correspond to stochastic versions with exponential probabilities.[2] This means that each pure strategy in each period is assigned an *attraction*, which is a function of the attraction in the previous period and the payoff to the particular strategy in the current period.[3] The attractions are then exponentially weighted in order to determine the mixed strategy to be employed in the next period. In the case of reinforcement learning, the attractions only depend on the payoff to the pure strategy actually chosen. In the case of belief learning, the hypothetical payoffs to the pure strategies that were not chosen are of equal importance (this is sometimes referred to as *hypothetical reinforcement*). However, Camerer and Ho's (1999) model also permits intermediate cases where payoffs corresponding to pure strategies that were not chosen are given a weight strictly between zero and one. The weight of such hypothetical payoffs is given by a single parameter, δ.

Camerer and Ho's (1999) model also allows initial attractions of different sizes which, in the case of belief learning, correspond to expected payoffs, given a prior distribution over the opponents' pure strategies. I depart from this assumption and set all initial attractions to zero, such that the individuals have almost no prior knowledge of the game they are drawn to play. This implies that the numerical analysis in this paper boils down to testing if any particular value δ corresponds to an ESLR.

In order to test the stability of learning rules, I simulate a large number of outcomes when all members of a finite population with a large share of incumbents and a small share of mutants are randomly matched for a finite number of periods. I then calculate the average payoff for each share of the popu-

[2] Fudenberg and Levine (1998) show that stochastic fictitious play can be derived by maximizing expected payoff given an empirical distribution of the opponents' past play when payoffs are subject to noise.

[3] The term "attraction" is used to make the terminology in this paper consistent with that in Camerer and Ho (1999). This term should not be interpreted in the mathematical sense, but as the weight assigned to a particular strategy.

lation. I consider four different games: Prisoners' Dilemma, Coordination, Hawk-Dove and Rock-Scissors-Paper. The main findings are:

- In almost all cases, the learning rule with full hypothetical reinforcement is an ESLR, whereas the learning rule with no hypothetical reinforcement is unstable.

- In the two games with no symmetric pure Nash equilibria – the Hawk-Dove and Rock-Scissors-Paper Games – the results depend on the level of *payoff sensitivity* of the learning rules. This is a parameter of the EWA model determining to what extent differences in attractions for the pure strategies should translate into differences in probabilities. For low payoff sensitivity, several rules appear to be stable, whereas for high payoff sensitivity, only belief learning is stable. The latter finding is, in part, due to that reinforcement learners with a high level of payoff sensitivity quickly become absorbed by a pure strategy, whereas belief learners with the same level of payoff sensitivity continue to adjust their mixed strategies until the last period.

- In 2×2 Coordination Games, the results also depend on the equilibrium payoffs. In such games, belief learning is generally a unique ESLR, but if the ratio of equilibrium payoffs becomes sufficiently small and payoff sensitivity is low, then there are also other stable learning rules.

1.1 Related Literature

The present paper is related to the theoretical literature on learning, but also to experimental tests of different learning rules. An early theoretical reference, asking similar questions, is Harley (1981). He analyzes the evolution of learning rules in the context of games with a unique evolutionarily stable strategy (ESS). He assumes the existence of an ESLR and then discusses the properties of such a rule. Harley claims that, given certain assumptions, "...the evolutionarily stable learning rule is a rule for learning evolutionarily

stable strategies." He also develops an approximation to such a rule and simulates its behavior in a homogeneous population. The current paper differs from that of Harley (1981) in that it explicitly formulates an evolutionary criterion for learning rules and does not assume the existence of an ESLR. Moreover, the analysis is not limited to games with a single ESS.

Anderlini and Sabourian (1995) develop a dynamic model of the evolution of algorithmic learning rules. They claim that under certain conditions, the frequencies of different learning rules in the population are globally stable and that the limit points of the distribution of strategies correspond to Nash equilibria. However, they do not investigate the properties of the stable learning rules.

Hopkins (2000) investigates the theoretical properties of stochastic fictitious play and perturbed reinforcement learning. The model in this paper is a special case of stochastic fictitious play, when the parameter δ is equal to one, and is similar to Hopkins' version of reinforcement learning when δ is equal to zero. Hopkins finds that the expected motion of both stochastic fictitious play and perturbed reinforcement learning can be written as a perturbed form of the replicator dynamics, and that in many cases, they will therefore have the same asymptotic behavior. In particular, he claims that they have identical local stability properties at mixed equilibria. He also finds that the main difference between the two learning rules is that fictitious play gives rise to faster learning. The analysis in Hopkins (2000) differs from the analysis in this paper, in that it is based on infinite interaction between two players using identical learning rules, but my findings are consistent with Hopkins' (2000) results.

The topic of this paper is also somewhat related to the theoretical literature on evolution in asset markets, such as Blume and Easley (1992, 2000), and Sandroni (2000). In these models, selection operates over beliefs and utility functions and not directly over learning rules, and the authors use a dynamic evolutionary criterion based on wealth accumulation. They find that, under

fairly general conditions, correct beliefs are selected for in complete markets, but not necessarily in incomplete markets.

The experimental literature uses a criterion which differs from the evolutionary one introduced in this paper to motivate the use of a particular learning rule. The objective is to find the learning rule which gives the best fit of experimental data. Camerer and Ho (1999) give a concise overview of the most important findings in earlier studies. They argue that the overall picture is unclear, but that comparisons appear to favor reinforcement in constant-sum games and belief learning in Coordination Games. In their own study of asymmetric Constant-Sum Games, Median-Action Games, and Beauty-Contest Games, they find support for a learning rule with parameter values in between reinforcement learning and belief based learning. In particular, they estimate game-specific values of the δ-parameter, which captures the degree of hypothetical reinforcement, strictly between zero and one, and generally around 0.5.

Stahl (2000) compares the prediction performance of seven learning models, including a restricted version of the EWA model. He pools data from a variety of symmetric two-player games and finds a logit best-reply model with inertia and adaptive expectations to perform best, closely followed by the EWA. For the latter, he estimates a value of the δ-parameter of 0.67.

This paper is organized as follows. Section 2 introduces the theoretical model underlying the simulations. Section 3 present the results of the Monte Carlo simulations. Section 4 contains a discussion of the results and Section 5 concludes. Tables and diagrams of some of the simulations can be found in the Appendix.

2 Model

Let Γ be a symmetric two-player game on normal form, where each player has a finite pure strategy set $X = \{x^1, ..., x^J\}$, with the mixed-strategy extension $\Delta(X) = \{p \in \Re_+^J \mid \sum_{j=1}^J p^j = 1\}$. Each player's payoff is represented by the

function $\pi : X \times X \to \mathbb{R}$, where $\pi(x, y)$ is the payoff to playing pure strategy x when the opponent is playing pure strategy y. From time to time, all individuals of a finite population, consisting of an even number M of individuals, are drawn to play this game for T periods. The mixed strategy of individual k in period $t \in \{1, 2, ..., T\} = \Upsilon$ is denoted by $p_k(t)$. The pure strategy realization of individual k is denoted by $x_k(t)$ and that of her opponent (in this period) by $y_k(t)$. The sequence

$$h_k(t) = \left((x_k(0), y_k(0)), (x_k(1), y_k(1)), ..., (x_k(t-1), y_k(t-1))\right),$$

where $(x_k(0), y_k(0)) = \emptyset$, is referred to as *individual k's history in period t*. Let $H(t)$ be the finite set of possible such histories at time t, let $H = \cup_{t=1}^{T} H(t)$, and let $\Omega = H(T)$ be the set of outcomes. I define a *learning rule* as a function $f : H \to \Delta(X)$ that maps histories to mixed strategies and denote the set of possible learning rules by $\mathfrak{F}$. Note that according to this definition, initial conditions such as initial history or initial strategy weights are given by the learning rule.

The matching procedure can be described as follows. In each of T periods, *all* members of the population are randomly matched in pairs to play the game Γ against each other. This can be illustrated by an urn with n balls, from which randomly selected pairs of balls (with equal probability) are drawn successively until the urn is empty. This procedure is repeated for a total of T periods, and the draws in each period are independent of the draws in all other periods. Each individual k receives a payoff $\pi(x_k(t), y_k(t))$ in each period and has a private history of realized strategy profiles. The expected payoff for an individual k employing learning rule f in a heterogenous population of size M, where the share of individuals employing rule f is $(1 - \varepsilon)$ and the share of individuals employing rule g is ε, is the expected average payoff under the probability measure, $\mu^M_{f,(1-\varepsilon)f+\varepsilon g}$ induced by the two rules present in the

population and their respective shares,

$$V^M(f,(1-\varepsilon)f+\varepsilon g) \tag{1}$$

$$=\sum_{h_k\in\Omega}\left(\frac{1}{T}\sum_{t=1}^{T}\pi(x_k(t),y_k(t))\right)\mu^M_{f,(1-\varepsilon)f+\varepsilon g}(h_k) \tag{2}$$

$$=E^M_{f,(1-\varepsilon)f+\varepsilon g}\left[\frac{1}{T}\sum_{t=1}^{T}\pi(x_k(t),y_k(t))\right], \tag{3}$$

where $\pi(x_k(t),y_k(t))$ refers to the realized payoff to individual k in period t, induced by history h_k.

Let $\mathfrak{F}'$ be an arbitrary non-empty subset of $\mathfrak{F}$. I define the following evolutionary stability criterion for learning rules.

Definition 1 *A learning rule $f\in\mathfrak{F}'$ is **evolutionarily stable** in the class $\mathfrak{F}'$ if for every $g\in\mathfrak{F}'\setminus f$, there exists an $\hat{\varepsilon}_g>0$ such that for all $\varepsilon\in(0,\hat{\varepsilon}_g)$,*

$$V^M(f,(1-\varepsilon)f+\varepsilon g)>V^M(g,(1-\varepsilon)f+\varepsilon g). \tag{4}$$

2.1 Experience Weighted Attraction Learning

In the present paper, I focus on a set of learning rules that can be described by Camerer and Ho's (1999) model of *experienced-weighted attraction* (EWA) learning. These are learning rules such that individual k's probability of strategy x^j in period $t\in\Upsilon$ can be written as

$$p^j_k(t)=\frac{e^{\lambda A^j_k(t-1)}}{\sum_{j=1}^{J}e^{\lambda A^j_k(t-1)}}, \tag{5}$$

where the attraction of strategy x^j is updated according to the formula

$$A^j_k(t)=\frac{\phi N(t-1)A^j_k(t-1)+\left[\delta+(1-\delta)I(x^j,x_k(t))\right]\pi(x^j,y_k(t))}{N(t)}, \tag{6}$$

for $t\in\Upsilon$, and $A^j_k(0)$ is a constant, and where

$$N(t)=\sigma N(t-1)+1, \tag{7}$$

for $t \in \Upsilon$, and $N(0)$ is a constant.[4] $I(x^j, x_k(t))$ is an indicator function which takes the value of one if $x_k(t) = x^j$ and zero otherwise, $y_k(t)$ is the realized pure strategy of the opponent in period t, and ϕ and σ are positive constants.

Note that this class of learning rules includes two of the most common learning rules used in the literature. When $\delta = 0$, $\sigma = \phi$ and $N(0) = \frac{1}{1-\sigma}$, EWA reduces to (average) reinforcement learning.[5] When $\delta = 1$, $\sigma = \phi$ and

$$A_k^j(0) = \sum_{l=1}^{J} \pi(x^j, y^l) \frac{N_{-k}^l(0)}{\sum_{l=1}^{J} N_{-k}^l(0)}, \tag{8}$$

where $\frac{N_{-k}^l(0)}{\sum_{l=1}^{J} N_{-k}^l(0)}$ is some initial relative frequency of strategy l, EWA becomes belief learning.

In order to make the analysis more tractable, I further restrict the set of rules to EWA learning rules such that $\sigma = \phi < 1$, $N(0) = \frac{1}{1-\sigma}$, and

$$A_k^j(0) = 0 \ \forall k, \forall j. \tag{9}$$

This means that the initial attractions will not generally correspond to those of belief learning. The assumption of equal initial attractions is motivated by a setting where the players have very limited information about the game before the first period and where they cannot use previous experience.[6] The assumption that $\sigma = \phi$ implies that the discount factor for a belief learner's historical observations of strategy realizations is the same as that for a reinforcement learner's historical attractions. Finally, the value of $N(0)$ corresponds to the steady state value of $N(t)$.[7]

[4] Camerer and Ho (1999) note that it is also possible to model probabilities as a power function of attractors.

[5] Camerer and Ho (1999) distinguishes between *average* and *cumulative* reinforcement, which results if $\rho = 0$ and $N(0) = 1$. The analysis in the present paper is based on average reinforcement.

[6] Although the game is fixed in the below analysis, a rationale for the assumption of uniform initial weight could be a setting where the game is drawn at random from some set of games before the first round of play.

[7] Stahl (2000) finds that a time varying $N(t)$ only improves the predictive power of the model marginally and assumes $N(t) = 1$ for all t. He also assumes all initial attractors to be zero and uses the updating formula to determine the attractors in period one.

I denote the set of rules with the above parameter values by $\mathfrak{F}_e$. Substituting in (6) and (7) gives

$$N(t) = \frac{1}{1-\sigma} \text{ for } t \in \Upsilon \tag{10}$$

and

$$A_k^j(t) = \sigma A_k^j(t-1) + (1-\sigma)\left[\delta + (1-\delta)I(x^j, x_k(t))\right]\pi(x^j, y_k(t)) \tag{11}$$

for $t \in \Upsilon$, and $A_k^j(0) = 0$. The formula in (5) now corresponds to belief learning (with modified initial weights) for $\delta = 1$ and to reinforcement learning for $\delta = 0$. The parameter δ captures the extent to which the hypothetical payoffs of pure strategies not played in a period are taken into account. σ is a constant determining the relative weights of recent and historical payoffs in the updating of mixed strategies.

3 Numerical Analysis

The analysis is based on Monte Carlo simulations of repeated encounters between individuals using different learning rules (i.e. with different values of δ) belonging to the set $\mathfrak{F}_e$. I focus on four types of games, Prisoners' Dilemma, 2×2 Coordination, Hawk-Dove, and Rock-Scissors-Paper Games. I generally set the payoff sensitivity parameter λ in (5) to either 1 or 10, σ equal to 0.95 and I assume that δ is an element of the set $D = \{0, 0.25, 0.5, 0.75, 1\}$, but I also test the robustness of my results by trying other parameter values (see the Appendix for a list of simulations).

In the simulations, each member of a population of 100 individuals, among which 10 are mutants with a different learning rule, is randomly matched with another member every period for $T = 100$ periods. The expected payoff to a learning rule is estimated by computing the mean of the average payoff for all individuals with the same learning rule in the population and by simulating 1000 such T-period outcomes. Since the mean payoff difference in each simulation is independently and identically distributed relative to the mean payoff difference in another simulation with the same population mixture,

the Central Limit Theorem applies and the mean payoff difference is approximately normally distributed. For each value of δ, the null hypothesis is that the corresponding learning rule is an ESLR. This hypothesis is rejected if the mean payoff to any mutant rule is statistically significantly higher than the mean payoff to the incumbent rule in the class, in accordance with Definition 1 above. More specifically, the decision rule is as follows. The null hypothesis,

$$H_0^\delta : f_\delta \text{ is an ESLR in the class } \mathfrak{F}_e,$$

is rejected in favor of the alternative hypothesis,

$$H_1^\delta : f_\delta \text{ is not an ESLR in the class } \mathfrak{F}_e,$$

if and only if, for some $\delta' \in D \setminus \delta$,

$$\frac{\hat{V}(f_\delta, (1-\varepsilon)f_\delta + \varepsilon f_{\delta'}) - \hat{V}(f_{\delta'}, (1-\varepsilon)f_\delta + \varepsilon f_{\delta'})}{\frac{1}{\sqrt{T}} s_\Delta \left[(1-\varepsilon)f_\delta + \varepsilon f_{\delta'}\right]} < -z_\alpha, \tag{12}$$

where $\hat{V}$ is the estimated average payoff, $s_\Delta \left[(1-\varepsilon)f_\delta + \varepsilon f_{\delta'}\right]$ denotes the sample standard deviation of the difference in mean average payoffs, computed over the 1000 simulations, and z_α is the critical value of the standard normal distribution.

3.0.1 Prisoners' Dilemma Games

Table 1 depicts the mean of the average payoffs among 90 incumbents, with a δ given in the left-most column, and 10 mutants, with a δ given in the top row, playing the game in Figure 1 for 1000×100 periods, when payoff sensitivity $\lambda = 10$. The value in brackets corresponds to the z-statistic of the differences in means, i.e. the difference, computed as the average incumbent payoff minus the average mutant payoff, divided by the standard error of the difference. As explained above, the null hypothesis that a learning rule with a particular δ is an ESLR can be rejected if this value is smaller than the critical value of the standard normal distribution, $-z_\alpha$ for some mutant learning rule in the class, different from f_δ.

It follows from the table that the null can be rejected for all learning rules except the one with $\delta = 1$ at the 10% and 5% significance level ($z_{0.05} = 1.645$, and $z_{0.10} = 1.282$). This is also illustrated by the diagram in Figure 2, where the standardized payoff difference (the z-statistic) between incumbent and mutant payoffs is plotted for different values of incumbent and mutant δ. In the diagram, the difference is set to zero for homogenous populations. The result is robust to changes in payoff sensitivity λ, initial conditions, payoff matrix, and the size of the mutant invasion.

The standard deviation of payoffs among learning rules with $\delta = 0$ is considerably larger than for other values of δ. The volatility of payoffs also depends on payoff sensitivity. If λ is reduced from 10 to 1, the range of standard deviations decreases considerably. For the high value of λ, convergence to the Nash equilibrium is fast. For the low value, the population share using the equilibrium strategy increases more slowly and keeps oscillating.

	y^1	y^2
x^1	4	0
x^2	5	3

FIGURE 1

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	2.8977 2.8876 (1.01)	2.8574 3.3005 (-70.20)	2.8570 3.3333 (-65.55)	2.8582 3.3303 (-65.97)	2.8572 3.3315 (-67.70)
0.25	3.0001 2.8610 (33.65)	2.9906 2.9907 (-0.14)	2.9891 3.0106 (-38.93)	2.9886 3.0166 (-54.25)	2.9883 3.0185 (-56.95)
0.50	3.0030 2.9018 (36.79)	2.9973 2.9788 (27.54)	2.9959 2.9962 (-0.88)	2.9955 3.0015 (-19.07)	2.9953 3.0031 (-26.46)
0.75	3.0039 2.9061 (38.71)	2.9987 2.9774 (33.98)	2.9975 2.9918 (14.86)	2.9970 2.9973 (-1.05)	2.9969 2.9988 (-6.85)
1.00	3.0037 2.9126 (39.39)	2.9991 2.9765 (34.83)	2.9979 2.9903 (21.62)	2.9975 2.9952 (8.07)	2.9973 2.9979 (-2.17)

TABLE 1–Mean payoffs and standardized payoff differences from playing the game in Figure 1.

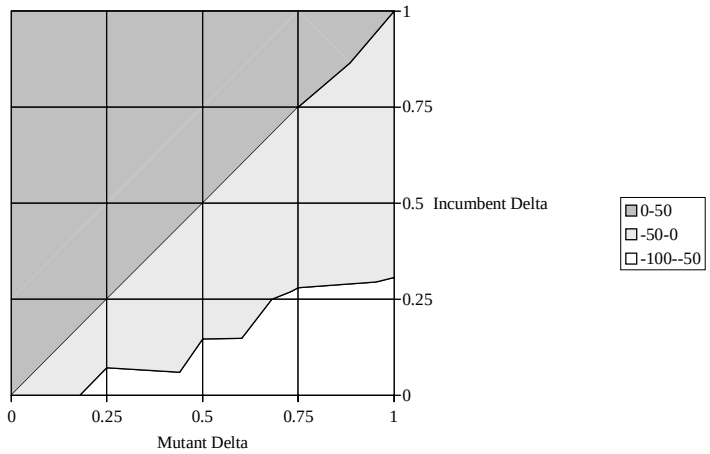

FIGURE 2–Standardized payoff difference between incumbents and mutants from Table 1.

3.0.2 2×2 Coordination Games

Table 2 and Figure 4 show the results from the simulations of the Coordination Game in Figure 3, with payoff sensitivity $\lambda = 10$. Once again, the null hypothesis can be rejected for all learning rules except $\delta = 1$ at 5% or 10% significance. This result is robust to changes in λ, the size of the invasion, and the initial conditions. However, when the ratio of diagonal payoffs is small ($\pi(x^1, y^1) = 1.1$ instead of 2) *and* $\lambda = 1$, then the null cannot be rejected for any of the rules $\delta = 1, \delta = 0.75$, and $\delta = 0.25$. From the table, it also follows that the outcome for a homogenous population of belief learners Pareto dominates that of a population of reinforcement learners.

	y^1	y^2
x^1	2	0
x^2	0	1

FIGURE 3

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.8583 1.8588 (-0.58)	1.8618 1.8770 (-24.61)	1.8640 1.8833 (-36.44)	1.8651 1.8866 (-40.05)	1.8654 1.8868 (-40.00)
0.25	1.8993 1.8866 (17.67)	1.9020 1.9020 (0.12)	1.9022 1.9069 (-11.17)	1.9042 1.9108 (-16.35)	1.9036 1.9113 (-19.32)
0.50	1.9131 1.8973 (23.86)	1.9148 1.9103 (10.45)	1.9157 1.9165 (-2.14)	1.9161 1.9184 (-6.12)	1.9168 1.9207 (-10.23)
0.75	1.9202 1.9022 (28.15)	1.9222 1.9152 (15.89)	1.9227 1.9206 (5.19)	1.9235 1.9239 (-1.17)	1.9240 1.9249 (-2.53)
1.00	1.9257 1.9074 (28.57)	1.9268 1.9194 (16.94)	1.9274 1.9235 (9.89)	1.9280 1.9260 (5.49)	1.9286 1.9290 (-1.28)

TABLE 2-Mean payoffs and standardized payoff differences from playing the game in Figure 3.

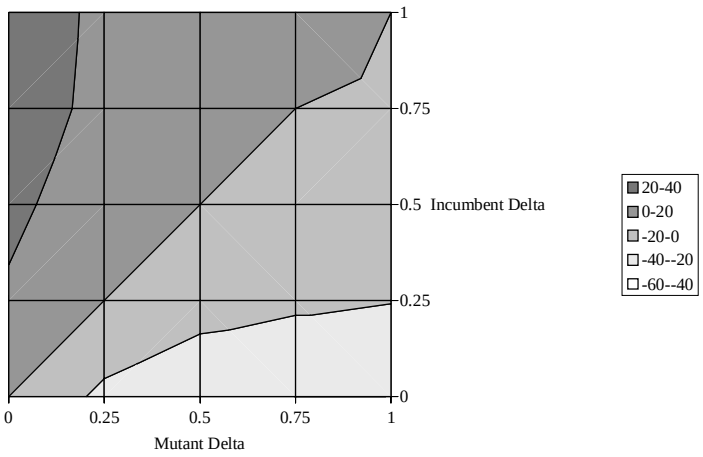

FIGURE 4–Standardized payoff difference between incumbents and mutants from Table 2.

3.0.3 Hawk-Dove Games

Table 3 and Figure 6 illustrate the results from the simulations of the game in Figure 5. For this game, the results are sensitive to the level of payoff sensitivity λ and the payoff matrix. When $\lambda = 10$, the null hypothesis cannot be rejected for the learning rules with $\delta = 0.75$ and $\delta = 1$ at the 10% level, whereas all other learning rules in the class appear to be unstable. If $\pi(x^1, y^2)$ is increased from 4 to 10, such that the initial uniform distribution is further

from the mixed equilibrium, then the null can be rejected for all rules except $\delta = 0.75$.

Part of the reason why learning rules with low δ are not stable and cannot invade other learning rules in this game when payoff sensitivity is high is that they quickly become absorbed by a pure strategy (see Figure 12 in the Appendix), something which can be exploited by learning rules with high δ that do not lock in on a particular pure strategy. This tendency for reinforcement learners to be absorbed has previously been noted by Fudenberg and Levine (1998).

For $\lambda = 1$, reinforcement learners no longer lock in on a particular strategy, but oscillate around the mixed equilibrium (see Figure 11 in the Appendix), which somewhat reduces the evolutionary advantage of belief-learners. For the game in Figure 5, the null hypothesis can be rejected for all rules except $\delta = 0.25$, $\delta = 0.75$, and $\delta = 1.0$ at the 10% level. In the game with $\pi(x^1, y^2) = 10$, the result is unchanged and all rules except $\delta = 0.75$ can be rejected.

The matrix in Table 3 also illustrates the potential trade-off between the Pareto efficiency and the evolutionary stability of a learning rule. The learning rule with $\delta = 0$ strictly dominates all other learning rules, but it is not sustainable since, in the case of an invasion, mutants with higher δ earn higher mean payoffs.

	y^1	y^2
x^1	0	4
x^2	1	2

FIGURE 5

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.7605 1.7612 (-0.20)	1.7508 1.7843 (-9.23)	1.7100 1.8353 (-36.42)	1.6692 1.8445 (-54.56)	1.6794 1.8397 (-50.94)
0.25	1.6889 1.6657 (8.06)	1.6819 1.6795 (0.82)	1.6518 1.7150 (-22.42)	1.6148 1.7302 (-43.28)	1.6238 1.7249 (-38.13)
0.50	1.5512 1.5206 (14.25)	1.5527 1.5283 (11.29)	1.5420 1.5402 (0.77)	1.5183 1.5505 (-14.54)	1.5256 1.5557 (-14.27)
0.75	1.4512 1.4279 (13.94)	1.4510 1.4306 (11.78)	1.4469 1.4346 (6.69)	1.4418 1.4402 (0.82)	1.4422 1.4397 (1.32)
1.00	1.4604 1.4391 (12.37)	1.4601 1.4397 (11.47)	1.4551 1.4419 (6.99)	1.4469 1.4473 (-0.19)	1.4468 1.4446 (1.21)

TABLE 3-Mean payoffs and standardized payoff differences from playing the game in Figure 5.

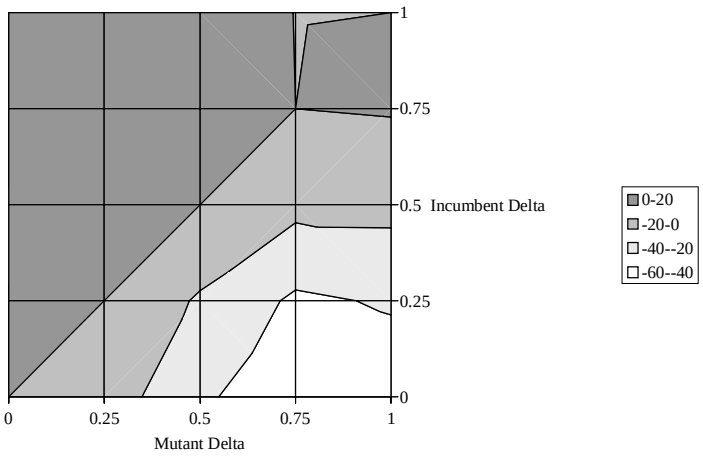

FIGURE 6–Standardized payoff difference between incumbents and mutants from Table 3.

3.0.4 Rock-Scissors-Paper

In the Rock-Scissors-Paper Game of Figure 7, the outcome is sensitive to payoff sensitivity. When the payoff sensitivity is $\lambda = 1$, the null hypothesis can be rejected for all learning rules in the class except $\delta = 0.0$, $\delta = 0.75$, and $\delta = 1.0$. All learning rules oscillate around the mixed equilibrium.

Table 4 and Figure 8 illustrate the case when $\lambda = 10$. As can be seen, the null hypothesis can be rejected for all learning rules except $\delta = 1$. As in the Haw-Dove game, the instability of rules with low δ for high values of payoff

sensitivity can, in part, be explained by their tendency to lock in on a pure strategy at an early stage.

	y^1	y^2	y^3
x^1	1	2	0
x^2	0	1	2
x^3	2	0	1

FIGURE 7

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.0001 0.9994 (0.58)	0.9987 1.0121 (-11.14)	0.9969 1.0278 (-23.92)	0.9970 1.0271 (-22.98)	0.9974 1.0236 (-21.35)
0.25	1.0006 0.9945 (5.46)	1.0000 0.9998 (0.25)	0.9984 1.0146 (-15.01)	0.9978 1.0200 (-21.19)	0.9981 1.0175 (-18.65)
0.50	1.0001 0.9988 (1.44)	1.0003 0.9976 (3.01)	0.9999 1.0009 (-1.16)	0.9995 1.0047 (-5.92)	0.9992 1.0071 (-8.89)
0.75	1.0004 0.9967 (4.16)	1.0002 0.9983 (2.24)	1.0002 0.9984 (2.22)	1.0000 1.0004 (-0.50)	0.9997 1.0026 (-3.34)
1.00	1.0003 0.9971 (3.66)	1.0003 0.9971 (3.59)	1.0005 0.9953 (6.06)	1.0002 0.9980 (2.65)	1.0000 0.9998 (0.28)

TABLE 4-Mean payoffs and standardized payoff differences from playing the game in Figure

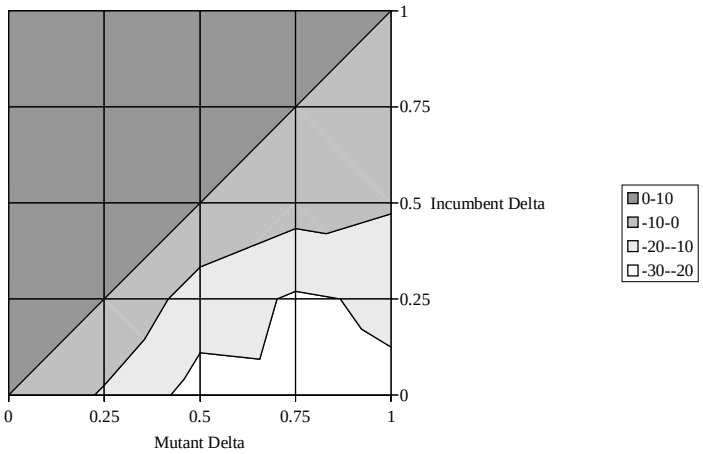

FIGURE 8–Standardized payoff difference between incumbents and mutants from Table 8.

3.1 Summary of Results

Table 9 summarizes the results of the simulations. The main finding is that belief learning is the only learning rule which is evolutionarily stable in almost all settings, whereas reinforcement learning is unstable in almost all settings.

In the Hawk-Dove and Rock-Scissors-Paper Games, the results depend on the payoff sensitivity. Learning rules with low degrees of hypothetical reinforcement are highly unstable for high payoff sensitivity. Part of the explanation is that such rules rapidly become absorbed by a pure strategy, whereas belief learners with the same level of payoff sensitivity continue to adjust their mixed strategies until the last period.

In Coordination and Hawk-Dove Games, the results also depend on the equilibrium payoffs. In the Coordination Game, belief learning is generally a unique ESLR, but if the ratio of equilibrium payoffs becomes sufficiently small, then there are also other learning rules for which the null cannot be rejected. Similarly, in the Hawk-Dove Game, belief learning with $\delta = 0.75$ is a unique ESLR for high payoff ratio, but for a smaller ratio it seems that there are also other stable learning rules.

Game	λ	**ESLR at the 10% significance level**
Prisoners' Dilemma	1	$\delta = 1.0$
	10	$\delta = 1.0$
Coordination	1	$\delta = 1.0$, for low payoff ratio $\delta = 0.25, 0.75, 1.0$
	10	$\delta = 1.0$, for low payoff ratio $\delta = 1.0$
Hawk-Dove	1	$\delta = 0.25, 0.75, 1.0$, for high payoff ratio $\delta = 0.75$
	10	$\delta = 0.75, 1.0$, for high payoff ratio $\delta = 0.75$
Rock-Scissors-Paper	1	$\delta = 0.0, 0.75, 1.0$
	10	$\delta = 1.0$

TABLE 9-Summary of the results from the different simulations.

4 Discussion

Hopkins (2000) investigates the theoretical properties of stochastic fictitious play and perturbed reinforcement learning in a setting where two individuals using identical learning rules interact for an infinite number of periods. He

demonstrates that the expected motion of both stochastic fictitious play and perturbed reinforcement learning can be written as a perturbed form of the replicator dynamics, and therefore, in many cases, will have the same asymptotic behavior. In particular, he claims that they have identical local stability properties at mixed equilibria and that the main difference between the two learning rules is that fictitious play gives rise to faster learning. The results in this paper indicate that speed of learning is indeed an important factor in explaining the stability of belief learning and that the difference between rules with high and low degrees of hypothetical reinforcement is smaller in games with mixed equilibria. However, other factors, such as a high probability of convergence to the equilibrium with the highest payoff in 2×2 Coordination Games and a low probability of absorption by a pure strategy in games with no symmetric pure equilibria, also seem important.

Camerer and Ho (2000) estimate separate sets of parameters for asymmetric Constant-Sum Games, Median-Action Games and Beauty-Contest Games. Their estimates of the degree of hypothetical enforcement, δ, are generally around 0.5, that of the discount factor, ϕ, in the range of 0.8 to 1.0, that of the second discount factor, σ, in the range of 0 to ϕ, and the payoff sensitivity, λ, varies from 0.2 to 18. Reinforcement learning and belief learning are generally rejected in favor of an intermediate model. Stahl (2000) pools data from several symmetric two-player games and estimates a δ of 0.67. Hence, the two studies lend support to the hypothesis that people take hypothetical payoffs into account, but especially the former study seems to find lower degrees of hypothetical reinforcement than predicted by the evolutionary analysis in this paper.

One should, however, be cautious in making a direct comparison with the results in Camerer and Ho (2000). First of all, the games played in their experiments differ considerably from, and are more complex, than the ones analyzed in this paper. Second, the learning rules in this paper do not exactly correspond to theirs. In particular, Camerer and Ho allow learning

rules with different initial attractions, whereas I assume that the players give equal weight to all their pure strategies at the start of the first period of play.

The setting in this paper is, at least in some respects, more similar to that in Stahl (2000). He also considers finite symmetric two-player games with and without symmetric pure equilibria. Moreover, he assumes the initial attractions of the EWA model to be zero, and use the updating formula to determine their values in period one.

A final comment concerns the environment where the learning rules operate. Although the game is fixed in this paper, the general idea is to find a learning rule which is evolutionarily stable under various conditions and can survive in a setting where the game changes from time to time – in many ways a more realistic description of human interaction. The results in this paper indicate that belief learning is indeed such a robust rule, although more analysis is needed to confirm this hypothesis.

5 Conclusion

In this paper, I define an evolutionary stability criterion for learning rules. I then apply this criterion to a class of rules which contains versions of two of the most well-known learning rules, reinforcement learning and belief learning, as well as intermediate rules in terms of hypothetical reinforcement. I perform Monte Carlo simulations of a matching scheme where all members of a large population are rematched in every period and I find that maximum or close to maximum hypothetical reinforcement is the only learning rule that is evolutionarily stable for almost all the games studied. I also find that evolutionary stability in some games hinges critically on payoff sensitivity and the relative payoffs of the game.

The objective of this paper is to take a step towards closing the open-endedness of the bounded rationality paradigm. A next step might be to apply this analysis to a larger set of learning rules or, more importantly, to obtain theoretical results which can explain the observations in this paper.

Appendix

Plots of Simulated Outcomes

The following diagrams illustrate the share of individuals playing strategy x^1 among 90 incumbents, using a learning rule with $\delta = 1$, and 10 mutants, using a learning rule with $\delta = 0$, in a single simulation. Initial attractions are zero for all pure strategies and $\sigma = 0.95$.

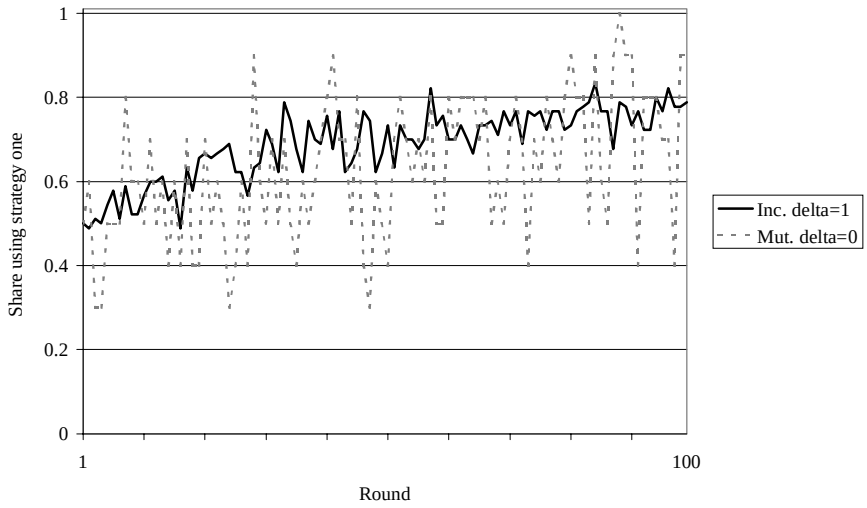

FIGURE 9–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in the game in Figure 3 when $\lambda = 1$.

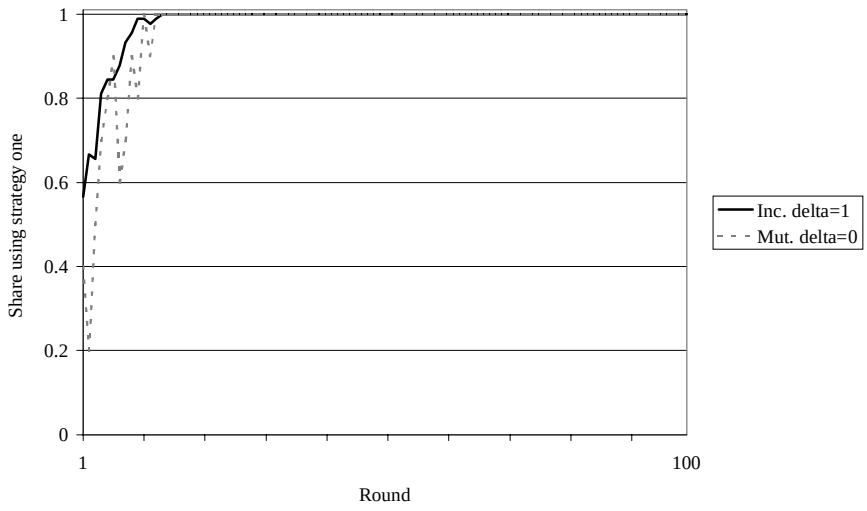

FIGURE 10–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in the game in Figure 3. when $\lambda = 10$.

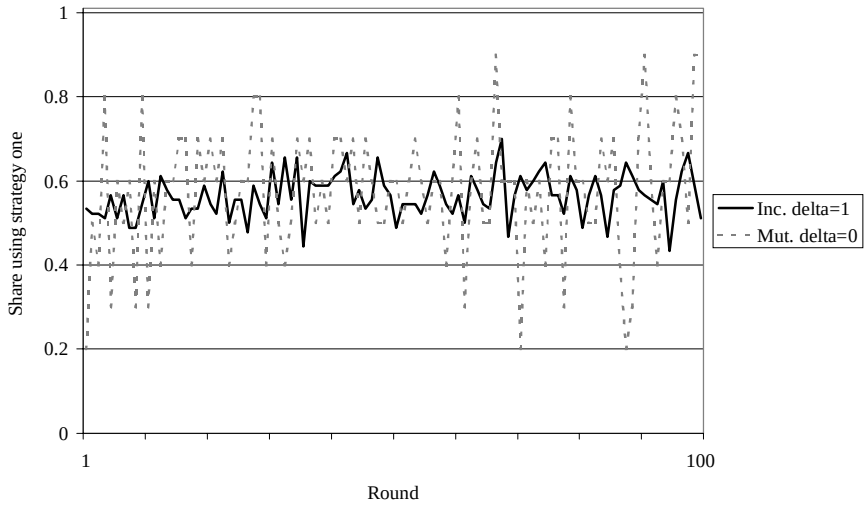

FIGURE 11–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in the game in Figure 5 when $\lambda = 1$.

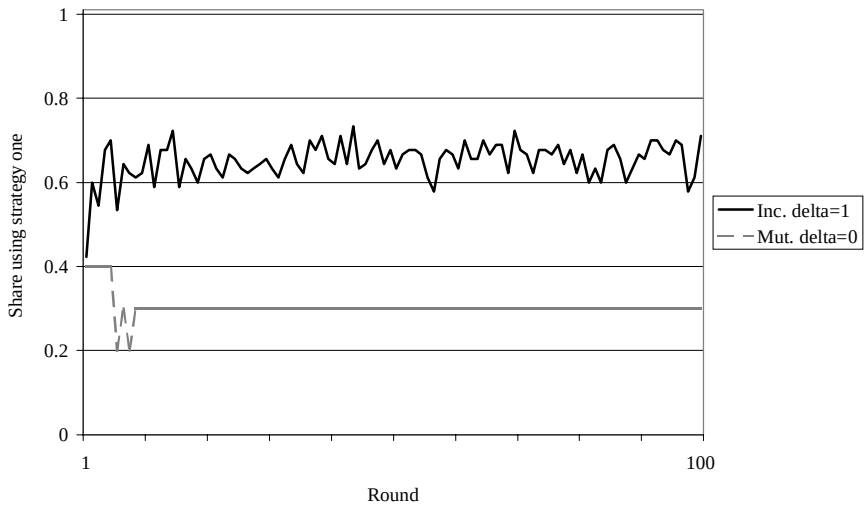

FIGURE 12–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in in the game in Figure 5 when $\lambda = 10$.

Tables of Simulated Payoffs

The following tables report the simulated average payoffs to an incumbent learning rule with a δ given by the left-most column, and an incumbent rule with a δ given by the top row. In each cell, the top-left number is the average payoff for all individuals with the incumbent rule, the top-right number the average payoff for all individuals with the mutant rule, and the figure in brackets is the difference between these numbers divided by the estimated standard error of the difference. In all simulations reported below, a population, consisting of 90 incumbents and 10 mutants, plays the game for 100 periods. Initial attractions are zero for all pure strategies and $\sigma = 0.95$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	2.9252 2.9239 (0.61)	2.9218 2.9576 (-21.86)	2.9201 2.9696 (-34.76)	2.9200 2.9722 (-39.55)	2.9195 2.9738 (-42.92)
0.25	2.9306 2.8955 (16.77)	2.9280 2.9241 (2.45)	2.9259 2.9425 (-11.59)	2.9252 2.9490 (-18.95)	2.9246 2.9543 (-24.89)
0.50	2.9328 2.8823 (24.80)	2.9298 2.9143 (9.96)	2.9282 2.9275 (0.49)	2.9270 2.9373 (-8.36)	2.9269 2.9410 (-12.20)
0.75	2.9342 2.8711 (33.24)	2.9312 2.9022 (18.88)	2.9296 2.9194 (7.86)	2.9286 2.9287 (-0.10)	2.9282 2.9341 (-5.23)
1.00	2.9346 2.8670 (36.11)	2.9321 2.8940 (24.95)	2.9304 2.9124 (13.73)	2.9295 2.9221 (6.10)	2.9286 2.9280 (0.61)

TABLE 10-Mean payoffs and standardized payoff differences from playing the game in Figure 1 when $\lambda = 1$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	0.8918 0.8906 (1.07)	0.8958 0.9064 (-9.08)	0.8992 0.9205 (-18.73)	0.9011 0.9293 (-25.21)	0.9015 0.9402 (-34.43)
0.25	0.9310 0.9172 (11.80)	0.9361 0.9341 (1.74)	0.9389 0.9496 (-9.50)	0.9421 0.9627 (-18.15)	0.9439 0.9713 (-25.32)
0.50	0.9692 0.9419 (24.00)	0.9734 0.9623 (10.10)	0.9765 0.9775 (-0.91)	0.9893 0.9896 (-0.26)	0.9838 1.0015 (-15.11)
0.75	1.0066 0.9657 (34.92)	1.0097 0.9868 (20.71)	1.0130 1.0023 (9.74)	1.0170 1.0169 (0.08)	1.0182 1.0269 (-7.83)
1.00	1.0390 0.9854 (45.31)	1.0441 1.0072 (30.63)	1.0464 1.0253 (17.68)	1.0488 1.0394 (8.12)	1.0527 1.0529 (-0.21)

TABLE 11-Mean payoffs and standardized payoff differences from playing the game in Figure 3 when $\lambda = 1$ and $\varepsilon = 0.1$.

	y^1	y^2
x^1	1.1	0
x^2	0	1

FIGURE 13

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	0.5263 0.5255 (1.45)	0.5262 0.5268 (-1.12)	0.5259 0.5270 (-1.99)	0.5258 0.5266 (-1.42)	0.5264 0.5265 (-0.13)
0.25	0.5262 0.5260 (0.33)	0.5261 0.5250 (2.16)	0.5263 0.5263 (-0.14)	0.5263 0.5257 (1.08)	0.5262 0.5265 (-0.64)
0.50	0.5265 0.5260 (0.90)	0.5264 0.5275 (-2.01)	0.5264 0.5259 (0.97)	0.5265 0.5274 (-1.63)	0.5270 0.5265 (0.87)
0.75	0.5269 0.5266 (0.60)	0.5268 0.5262 (1.15)	0.5274 0.5278 (-0.59)	0.5272 0.5262 (1.73)	0.5274 0.5273 (0.17)
1.00	0.5269 0.5263 (1.02)	0.5271 0.5276 (-1.02)	0.5271 0.5274 (-0.52)	0.5266 0.5261 (0.87)	0.5274 0.5270 (0.78)

TABLE 12-Mean payoffs and standardized payoff differences from playing the game in Figure 13 when $\lambda = 1$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	0.7548 0.7543 (0.48)	0.7686 0.7980 (-34.63)	0.7817 0.8181 (-41.35)	0.7738 0.8117 (-45.47)	0.7757 0.8131 (-46.57)
0.25	0.8888 0.8602 (26.82)	0.8938 0.8936 (0.33)	0.9067 0.9147 (-15.91)	0.9054 0.9159 (-22.81)	0.9054 0.9158 (-22.31)
0.50	0.9302 0.8950 (32.96)	0.9376 0.9308 (12.82)	0.9446 0.9451 (-1.35)	0.9431 0.9457 (-6.86)	0.9415 0.9441 (-6.64)
0.75	0.9473 0.9127 (32.57)	0.9581 0.9493 (17.28)	0.9562 0.9538 (5.90)	0.9632 0.9638 (-1.66)	0.9626 0.9634 (-2.41)
1.00	0.9604 0.9279 (32.69)	0.9656 0.9546 (22.00)	0.9713 0.9680 (8.64)	0.9718 0.9709 (2.56)	0.9727 0.9728 (-0.43)

TABLE 13-Mean payoffs and standardized payoff differences from playing the game in Figure 13 when $\lambda = 10$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.6079 1.6072 (0.35)	1.6067 1.6074 (-0.37)	1.6082 1.6094 (-0.67)	1.6080 1.6104 (-1.43)	1.6080 1.6111 (-1.80)
0.25	1.6055 1.6053 (0.07)	1.6053 1.6057 (-0.22)	1.6050 1.6061 (-0.62)	1.6060 1.6055 (0.29)	1.6066 1.6057 (0.52)
0.50	1.6073 1.6062 (0.64)	1.6060 1.6076 (-0.90)	1.6067 1.6041 (1.54)	1.6066 1.6079 (-0.73)	1.6067 1.6089 (-1.40)
0.75	1.6097 1.6103 (-0.33)	1.6089 1.6099 (-0.57)	1.6103 1.6070 (1.89)	1.6100 1.6115 (-0.88)	1.6100 1.6088 (0.73)
1.00	1.6124 1.6095 (1.63)	1.6117 1.6134 (-1.00)	1.6123 1.6132 (-0.52)	1.6123 1.6125 (-0.01)	1.6132 1.6124 (0.46)

TABLE 14-Mean payoffs and standardized payoff differences from playing the game in Figure 5 when $\lambda = 1$ and $\varepsilon = 0.1$.

	y^1	y^2
x^1	0	10
x^2	1	2

FIGURE 14

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	2.2135 2.2164 (-0.57)	2.2103 2.2357 (-5.10)	2.2131 2.2311 (-3.71)	2.2150 2.2248 (-2.08)	2.2188 2.2322 (-2.95)
0.25	2.2072 2.1882 (3.51)	2.2026 2.2008 (0.35)	2.2042 2.2062 (-0.41)	2.2044 2.2130 (-1.88)	2.2082 2.2052 (0.63)
0.50	2.2166 2.1936 (4.34)	2.2134 2.2101 (0.66)	2.2130 2.2124 (0.14)	2.2160 2.2138 (0.46)	2.2181 2.2105 (1.65)
0.75	2.2349 2.2071 (5.30)	2.2300 2.2271 (0.56)	2.2290 2.2365 (-1.55)	2.2342 2.2262 (1.65)	2.2362 2.2219 (3.10)
1.00	2.2524 2.2383 (2.72)	2.2485 2.2524 (-0.81)	2.2476 2.2550 (-1.49)	2.2502 2.2627 (-2.72)	2.2542 2.2553 (-0.23)

TABLE 15-Mean payoffs and standardized payoff differences from playing the game in Figure 14 when $\lambda = 1$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	3.1040 3.0822 (1.34)	2.8984 3.8830 (-83.83)	2.8590 3.9861 (-112.93)	2.8618 4.0014 (-118.06)	2.8655 4.0164 (-119.27)
0.25	2.2468 1.9268 (39.42)	2.1574 2.1666 (-1.32)	2.0830 2.2953 (-37.73)	2.0814 2.3095 (-43.77)	2.0944 2.3237 (-42.27)
0.50	1.7555 1.5211 (47.15)	1.7190 1.6076 (22.07)	1.6824 1.6874 (-1.09)	1.6824 1.6896 (-1.62)	1.6872 1.6863 (0.20)
0.75	1.7327 1.4996 (48.20)	1.7002 1.5665 (28.17)	1.6563 1.6369 (4.29)	1.6492 1.6517 (-0.57)	1.6537 1.6473 (1.52)
1.00	1.7903 1.5412 (44.56)	1.7503 1.6142 (24.55)	1.6979 1.6945 (0.73)	1.6903 1.7068 (-3.69)	1.6966 1.6968 (-0.04)

TABLE 16-Mean payoffs and standardized payoff differences from playing the game in Figure 14 when $\lambda = 10$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.0000 1.0002 (-0.22)	1.0001 0.9992 (0.98)	1.0000 1.0002 (-0.19)	1.0000 1.0000 (0.00)	1.0002 0.9983 (2.14)
0.25	0.9998 1.0016 (-2.01)	1.0000 0.9996 (0.57)	1.0000 1.0000 (-0.03)	0.9998 1.0020 (-2.66)	1.0000 1.0004 (-0.53)
0.50	1.0000 0.9999 (0.11)	1.0000 0.9997 (0.44)	1.0000 0.9999 (0.07)	0.9998 1.0019 (-2.39)	1.0000 1.0001 (-0.08)
0.75	1.0001 0.9995 (0.64)	1.0001 0.9988 (1.52)	1.0000 1.0001 (-0.11)	1.0001 0.9994 (0.79)	1.0001 0.9992 (1.14)
1.00	1.0001 0.9995 (0.64)	0.9999 1.0005 (-0.67)	1.0000 1.0000 (0.01)	1.0000 1.0001 (-0.13)	1.0000 0.9998 (0.27)

TABLE 17-Mean payoffs and standardized payoff differences from playing the game in Figure 7 when $\lambda = 1$ and $\varepsilon = 0.1$.

References

[1] Anderlini, L., and H. Sabourian (1995): "The Evolution of Algorithmic Learning Rules: A Global Stability Result," Mimeo, Cambridge University.

[2] Blume, L., and D. Easley (1992): "Evolution and Market Behavior," *Journal of Economic Theory*, 58(1), 9-40.

[3] ——— (2000): "If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets," Mimeo, Cornell University.

[4] Bush, R. and F. Mosteller (1951): "A Mathematical Model for Simple Learning," *Psychological Review* 58, 313-323.

[5] Camerer, C. and T-H. Ho (1999): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, 827-874.

[6] Fudenberg, D., and D. Levine (1998): *The Theory of Learning in Games.* The MIT Press.

[7] Harley, C. B. (1981): "Learning the Evolutionarily Stable Strategy," *Journal of Theoretical Biology*, 89, 611-633.

[8] Hopkins, E. (2000): "Two Competing Models of How People Learn in Games," Mimeo, University of Edinburgh.

[9] Maynard Smith, J., and G. R. Price (1973): "The Logic of Animal Conflict," *Nature*, 246, 15-18.

[10] Maynard Smith, J. (1974): "The Theory of Games and the Evolution of Animal Conflicts," *Journal of Theoretical Biology*, 47, 209-221.

[11] Sandroni, I. (2000): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 68, 1303-1341.

[12] Stahl, D. O. (2000): "Action-Reinforcement Learning Versus Rule Learning," Mimeo, University of Texas.

[13] Thorndike, E. L. (1898): "Animal Intelligence: An Experimental Study of the Associative Process in Animals," *Psychological Monographs* 2.