Stochastic Stability in Discounted Stochastic Fictitious Play

Noah Williams*

Department of Economics, University of Wisconsin - Madison
E-mail: nmwilliams@wisc.edu

Revised January 18, 2018

In this paper I study how a model of stochastic fictitious play gives rise to switches between equilibria similar to stochastic evolutionary models, and characterize the long run behavior of the game. I focus on a model in which agents' payoffs are subject to random shocks and they discount past observations exponentially. I analyze the behavior of agents' beliefs as the discount rate on past information becomes small but the payoff shock variance remains fixed. I show that agents tend to be drawn toward an equilibrium, but occasionally the stochastic shocks lead agents to endogenously shift between equilibria. I then calculate the invariant distribution of players' beliefs, and use it to determine the most likely outcome observed in long run. Our application shows that by making some slight changes to a standard learning model, I can derive an equilibrium selection criterion similar to stochastic evolutionary models but with some important differences.

1. INTRODUCTION

Numerous economic models have multiple equilibria, which immediately raises the question of how to characterize outcomes or to select among equilibria. In this paper I develop methods to characterize the long run behavior of discrete time models with multiple stable equilibria. This allows me to characterize the distribution over equilibria and to determine which among many possible equilibria is most likely to be observed in the long run. This is a problem which has long been studied in evolutionary game theory, but here I show how equilibrium selection may result from a model of individual agent learning.

This paper follows much of the recent literature in viewing the equilibrium of a game as resulting from a process of learning by agents. Since many games have multiple Nash equilibria, a common question is which equilibria would be likely outcomes of an adaptive process. For example, models of learning have sought to determine which equilibria are stable under a specified learning process. However, this criterion is not sufficient to determine which equilibrium is the most plausible in models where there are multiple stable equilibria. In contrast, evolutionary models starting with the seminal work of Foster and Young (1990), Kandori, Mailath, and Rob (1993) and Young (1993) among others, have sought to determine which equilibria are stable

* I thank Xiaohong Chen, In-Koo Cho, Rama Cont, Xavier Gabaix, Lars Peter Hansen, Ken Kasa, David Kreps, Markus Mobius, William Sandholm, Thomas Sargent, Peyton Young and especially Dana Heller and Glenn Ellison for helpful comments, discussions, and suggestions.

outcomes of a long run process of evolution in which occasional random shocks perturb agents' decisions. These models have provided sharp characterizations of the most likely equilibrium observed in the long run, which is often unique, although it typically depends on the details of the adjustment and perturbation process.

This paper provides a partial synthesis of the lines of research on learning and long run equilibrium selection. I analyze the well-known stochastic fictitious play learning model in which agents' payoffs are subject to random shocks. I study a variant of the model in which agents learn by placing a constant weight or "gain" on new information, and so discount past observations exponentially. I analyze the behavior of agents' beliefs as this discount rate gets small, but the shock variance is bounded away from zero. I provide conditions which insure that agents' beliefs converge to a stable equilibrium in a distributional sense. However the persistence of randomness and the nonvanishing weight on new information leads agents to continually revise their beliefs, and leads to occasional transitions between equilibria. This allows me to determine a stationary distribution of agents' beliefs which asymptotically places all of its mass on a unique long run equilibrium.

A main contribution of this paper is the development of applicable methods to analyze the long run behavior of discrete-time continuous-state models with multiple equilibria. In particular, I use techniques from large deviation theory to analyze the rare events in which agents escape from a stable equilibrium. These methods allow me to characterize transitions between equilibria by solving deterministic dynamic control problems. I discuss below some related known results for continuous time models and models with discrete state spaces. However, as many economic models are naturally cast in discrete time and have continuous state spaces, our results may be more broadly applicable beyond the specific model we consider here. Toward this end, I formulate the key theoretical results in general terms, and then show how they specialize in the case of fictitious play.

Similar large deviation methods have been recently studied by Sandholm and Staudigl (2016) who study a closely related dynamic. They focus on large population and small noise limits of a noisy best response model where agents are matched from finite population. Their general approach is similar to mine, but the analysis differs. I focus on games with a fixed number of players and fixed noise, taking a limit in the (fixed) gain in the learning algorithm. In Williams (2018) I take a similar approach in models with a unique equilibrium.

My model of learning in games is a variation on the stochastic fictitious play model which was introduced and first analyzed by Fudenberg and Kreps (1993). Stochastic fictitious play (SFP) introduces random shocks to players' payoffs in the spirit of the purification results of Harsanyi (1973) to the original (deterministic) fictitious play model of Brown (1951) and Robinson (1951). In the first part of the paper I consider the stability of equilibria under learning. In previous analyses of this model, Fudenberg and Kreps (1993) showed that in games with a unique Nash equilibrium in mixed strategies, play under this learning scheme converges to the Nash equilibrium. Kaniovski and Young (1995) extended these results to a general class of 2×2 games, and Benaim and Hirsch (1999a) further determined convergence criteria for a class of two action games with N players, with the most general results provided by Hofbauer and Sandholm (2002). Some related results have been shown by Hofbauer and Hopkins (2000) and Benaim (2000) who provide conditions for global convergence in certain classes of games. I summarize and apply these stability results to our discounted stochastic fictitious play model for two-player multi-action games.[1] As an example of stability analysis, I prove a conjecture by Fudenberg and Kreps (1993) about the stability of the Shapley (1964) game, showing that the unique Nash equilibrium is unstable for small noise.

After establishing stability of equilibria, I turn to the long run behavior of the adaptive system induced by the players' beliefs and actions. Under constant gain learning the weight on current observations is always nonzero, and thus the ongoing exogenous shocks insure that there is persistent randomness in the system. Although agents' beliefs converge to a stable equilibrium in a distributional sense, occasional sequences of shocks lead agents to change their strategy choices and can induce occasional "escapes" from a stable equilibrium. I formulate a deterministic control problem that provides the (probabilistic) rate of transition between equilibria. Following Freidlin and Wentzell (1999), I then calculate the stationary distribution of beliefs over equilibria, and show that typically this distribution is asymptotically concentrated on a unique equilibrium. Thus as time evolves the system will tend to spend most of its time within a neighborhood of a particular equilibrium, which in the literature following Kandori, Mailath, and Rob (1993) and Young (1993) has been called the long run or stochastically stable equilibrium. For general games, I provide expressions which must be evaluated numerically in order to determine the stochastically stable run equilibrium. However for the important special case of 2×2 symmetric games, I establish that the long run equilibrium is the risk dominant equilibrium. As I discuss below, this result agrees with many in the literature. However for larger games my results differ from existing criteria, as I show in an example below.

In addition to the papers already discussed, there is a long literature with similar aims and approaches. As noted above, there are related results for discrete-state-space models such as Kandori, Mailath, and Rob (1993) and especially Young (1993). These papers consider the dynamics in which "mutations" or mistakes perturb agents' choices and use arguments similar to those in this paper to characterize stochastic stability. Here I consider a discrete time model with a continuous state space.[2] Aside from this technical difference, there is a difference in focus. Rather than perturbing agents' decisions directly, I assume that there are shocks to agents' payoffs.[3] This

[1]In a related model, Ellison and Fudenberg (2000) considered the local stability of equilibria in 3×3 games. Their notion of purification differs from our specification of stochastic shocks, and their development is in continuous time.

[2]Foster and Young (1990) and Fudenberg and Harris (1992) present continuous-time continuous-state evolutionary models which use techniques similar to this paper.

[3]Myatt and Wallace (2004) also consider equilibrium selection in a 2×2 model with stochastic payoff shocks. They consider a different adjustment dynamic in which each period only one agent drawn from a population can revise his strategy. As discussed above, Sandholm and Staudigl (2016) also consider perturbed payoffs.

exogenous randomness interacts with agents' learning rules, and may lead them to switch strategies. The fact that agents' choice is directed, instead of being completely random, changes the probabilities that low payoff actions are played. This leads to some differences in long run equilibrium selection. In addition, most of the results in literature considered limits as the stochastic perturbations decrease to zero. In my model, the environment remains fully stochastic in the limit which may be more natural for many economic models which are fundamentally stochastic. In formulating this type of limit, my results are broadly similar to Binmore, Samuelson, and Vaughn (1995) and Boylan (1995). In addition to Sandholm and Staudigl, the paper closest to my results, although in a somewhat different setting, is Benaim and Weibull (2003). Our large deviation results are formally similar, but their paper analyzes large population limits and considers a rather different adjustment process.

Related studies of convergence, but not stochastic stability, have recently been based on models with stochastic choice using weaker informational requirements. For example, Leslie and Collins (2006) study convergence in a model where beliefs are subject to a general perturbation sequence, with stochastic fictitious play as a special case. They also study a simpler payoff-based approach, where agents only observe their own payoffs and not their opponents' actions. Cominetti, Melo, and Sorin (2010) study a related payoff-based process where agents only form estimates of payoffs of each action and make perturbed choice based on their estimates. Bravo (2016) modifies this approach to weight more heavily actions which have been played more frequently, which is similar to our discounting of past observations.

The rest of the paper is organized as follows. In the next section I present the model of stochastic fictitious play. In Section 3, I analyze the convergence of the beliefs, applying results from stochastic approximation theory and analyzing some examples fictitious play. In Section 4, I turn to analyzing the escape problem and characterize stochastic stability in a general framework. First I present the results from large deviation theory which allow one to compute the rates of escape from the different equilibria. These results provide a characterization of the stationary distribution of beliefs, and therefore determine the long run equilibrium. In Section 5, I analyze the stochastic fictitious play model and determine the stochastically stable equilibrium in some example games. Section 6 concludes. Technical assumptions and proofs of some results are collected in Appendix A.

2. DISCOUNTED STOCHASTIC FICTITIOUS PLAY

In this section I briefly present the model of discounted stochastic fictitious play (SFP). Under discounted SFP, instead of averaging evenly over the past observation of their opponent's play, agents discount past observations and put more weight on more recent ones. Throughout we restrict our attention to two player games in which the payoffs to each player are subject to stochastic shocks.[4] The game is repeated

[4]Hofbauer and Sandholm (2002) show that the same dynamics can result with deterministic perturbations. However the stochastic nature is important for our characterization of long run equilibria.

a possibly infinite number of times, but at each round each player treats the game as static and myopically chooses a pure strategy best response. (The only dynamics come through the evolution of beliefs.) The assumption of myopia can be motivated either by assuming bounded rationality or as a result of random matching of players from a large population.

For simplicity I focus on two player games where each player has the choice of N actions. Extensions to games with differing action spaces is straightforward, and the techniques could be adapted to more players as well. Before an agent decides which action to select, he observes a stochastic shock to his payoffs that is not observable to his opponent. Formally, we assume player 1's payoffs are $a_{i,j} + e^1_{i,t}$ when he plays i and player 2 plays j, for $i, j = 1, .., N$. Here $a_{i,j}$ represents the mean payoff and $e^1_{i,t}$ is a mean zero random variable which is common to the player's action (note that it does not depend on j.) Analogously, player 2's payoffs are given by $b_{i,j} + e^2_{i,t}$, so that the payoff bi-matrix has entries $(a_{i,j} + e^1_{i,t}, b_{j,i} + e^2_{j,t})$. Player 1 assesses probability $\theta_{2,i,t}$ that player 2 plays action i at date t, with $\theta_{1,i,t}$ defined analogously. Define the $(N \times 1)$ vectors $a_i = (a_{i,1}, ..., a_{i,N})'$ and $\theta_{2,t} = (\theta_{2,1,t}, ..., \theta_{2,N,t})'$, again with the obvious analogues b_j and $\theta_{1,t}$. For simplicity we assume:

Assumption 2.1. *The shocks $e^i_{j,t}$ have the common continuous distribution function F, and are independent across actions, across agents, and over time:* $e^i_{1,t} \perp e^i_{2,t}, e^1_{j,t} \perp e^2_{j,t}, e^i_{j,t} \perp e^i_{j,s}$, *for* $i, j = 1, 2,\ t, s > 0,\ i \neq j,\ t \neq s$.

We focus on the two special cases in which the errors are normally distributed with mean zero and variance σ^2 so that $F(x) = \Phi(\frac{x}{\sigma})$, and when the errors have a type-II extreme value distribution with parameter λ, in which case $F(x) = \exp(-\exp(-\lambda x - \gamma))$. Here γ is the Euler-Mascheroni constant which insures that the mean is zero. Note that as $\sigma \to 0$ and $\lambda \to \infty$ the shock distributions become more concentrated around zero. As we will see, these two distributions give rise to probit and logit decision rules, respectively. Some of our results hold for more general shock distributions, but these cases are the most commonly used and they allow us to obtain explicit results.

At each date, each player plays a myopic pure strategy best response based on his current beliefs, and then updates his beliefs about the other player's behavior based on his observations. Since the shocks have continuous distributions, there is (almost surely) no loss in generality in considering only pure strategies, and so we use "strategy" and "action" synonymously throughout. Thus, at date t, player 1 chooses action i if it yields the highest subjectively expected payoff:

$$\theta_{2,t} \cdot a_i + e^1_{i,t} \geq \max_{j \neq i} \left\{ \theta_{2,t} \cdot a_j + e^1_{j,t} \right\}. \tag{1}$$

(Any tie-breaking rule will suffice for the zero probability event that there are multiple maximal actions.) Then each player observes the opponent's action, and the players' beliefs are updated according to the following learning rule:

$$\begin{aligned}\theta_{1,i,t+1} &= \theta_{1,i,t} + \varepsilon\left[1_{\{\text{Player 1 plays } i\}} - \theta_{1,i,t}\right] \\ \theta_{2,j,t+1} &= \theta_{2,j,t} + \varepsilon\left[1_{\{\text{Player 2 plays } j\}} - \theta_{2,j,t}\right],\end{aligned} \tag{2}$$

where $1_{\{x\}}$ is an indicator function for the outcome x and ε is the "gain" which determines the weight on current observations relative to the past. Our discounted SFP model assumes that ε is a constant, and hence it is known as a "constant gain" algorithm.

Previous analyses of stochastic fictitious play have focused on the case were ε decreases over time as $1/t$.[5] With this gain setting, the learning algorithm is just a procedure for recursively estimating the empirical distribution of the opponent's strategies. Underlying this specification of the learning rule is the assumption that each agent believes that his opponent is drawing his strategy choices from a fixed distribution. As an agent gains more observations, he refines his estimates of this fixed underlying distribution. Since he considers the opponent's strategy distribution to be time invariant, all draws from this distribution are weighted equally.

In the discounted case, the gain is constant, and each player recognizes the possibility that the other player's beliefs may change over time, so observations are discounted at an exponential rate. This implies that recent observations are given more weight in estimation. This seems reasonable in our model because both agents are learning, so that their strategies are not drawn from a fixed distribution. Due to this nonstationarity of the system, a discounted algorithm may be more appropriate. The particular discounted rule is only optimal in some special cases, but it is clear that in nonstationary environments, discounted rules may outperform rules which weight observations equally. Moreover, a discounted specification was used by Cheung and Friedman (1997) in their empirical analysis of experimental data on learning in games. Their typical estimated discount rates were much less than one, and thus were consistent with relatively large settings of the gain ε. In the context of a different learning model, Sarin and Vahid (1999) also used a similar discounted specification under the assumption that agents did not know whether the environment was stationary. By placing a constant weight on current observations relative to the past, the discounted learning rule allows beliefs to react to the persistent randomness in the system, and this leads to the characterization of stochastic stability.

3. STABILITY AND CONVERGENCE

In this section, I begin analysis of the agents' beliefs by characterizing the sense in which beliefs converge and identifying the limit sets. First, I state some general convergence results from stochastic approximation theory, due to Kushner and Yin (1997) and Benaim (1999), which are relevant for the current model. In particular, I show that the limiting behavior of the learning rule is governed by a differential

[5]However, as was pointed out by Ellison and Fudenberg (2000), the stability conditions are identical in this case and in our discounted model.

equation. I then summarize known results on convergence, and consider two example games.

3.1. Convergence Results

In this section I summarize the existing results on the convergence of our discounted stochastic fictitious play learning rule. I apply results on two action games due to Fudenberg and Kreps (1993) for games with a unique mixed equilibrium, and Kaniovski and Young (1995) and Benaim and Hirsch (1999a) for the general case. I also include some special cases of larger games due to Hofbauer and Hopkins (2000) for zero sum and "partnership" games, and Benaim and Hirsch (1999b) and Benaim (2000) for "cooperative" games.

I first rewrite the learning rule above in a more abstract form. First I stack the beliefs of both players into a vector θ, all the shocks into a vector e and the "update" terms into a function b. Then we can write the learning rule as:

$$\theta_{t+1} = \theta_t + \varepsilon b(\theta_t, e_t). \tag{3}$$

Assumption A1 above restricts the error process e_t to be i.i.d. This could be weakened to allow for some forms of temporal dependence at a cost of complexity. I find it useful in the analysis to split b into its expected and martingale difference components:

$$\begin{aligned} \bar{b}\left(\theta_t\right) &= Eb(\theta_t, e_t), \\ v_t &= b(\theta_t, e_t) - \bar{b}\left(\theta_t\right). \end{aligned}$$

Thus we have the alternate form of (3):

$$\theta_{t+1} = \theta_t + \varepsilon \bar{b}\left(\theta_t\right) + \varepsilon v_t. \tag{4}$$

In Appendix A.1 I provide details on the calculation of $\bar{b}$ in my setting.

The convergence theorems below show that the limit behavior of (4) can be characterized by a differential equation. I now provide some heuristic motivation for the results. Note that we can re-write (4) as:

$$\frac{\theta_{t+1} - \theta_t}{\varepsilon} = \bar{b}\left(\theta_t\right) + v_t. \tag{5}$$

On the left side of (5), we have the difference between consecutive estimates, normalized by the gain. We can then think about embedding the discrete time process onto a continuous time scale, interpolating between the discrete observations and letting ε be the time between observations. Thus the estimates are θ_t and θ_{t+1} are ε units of time apart, so that the left side of (5) is a finite-difference approximation of a time derivative. As $\varepsilon \to 0$, this approximation will converge to the true time derivative. Turning to the right side of (5), the first term is a constant function of θ_t, while the second term is the martingale difference between the new information in the latest observation and its expectation. For small gain ε, agents average more evenly over the past, and so this difference is likely to be small. In particular, below we apply a law of large numbers to ensure that $v_t \to 0$ as $\varepsilon \to 0$. Together, these results imply that as $\varepsilon \to 0$, the dynamics of agents' beliefs in (4) converge to the trajectories of the differential equation:

$$\dot{\theta} = \bar{b}(\theta). \tag{6}$$

The same ODE characterizes the limits of the standard, un-discounted SFP learning rule which has been studied in the literature. However the asymptotic results are different. In the standard case, the gain is not constant but shrinks to zero as $t \to \infty$. Thus we take the limit *along a sequence of iterations*, and so for large t we can approximate the behavior of beliefs by the differential equation (6). In the discounted case, the gain is fixed along a given sequence as t increases. Therefore we look *across sequences of iterations*, each of which is indexed by a strictly positive gain. To emphasize this dependence I now add the superscript and denote the belief vector θ_t^ε. These different limits also lead to different convergence notions. In the usual case, beliefs typically converge with probability one, while in the discounted case they only converge weakly.

The basis for our convergence results are provided by the following theorem condensed from results of Kushner and Yin (1997) and Benaim (1999). In Appendix A.2 we list the necessary assumptions for the theorem, along with more details about the continuous time interpolation leading to (6). The theorem ensures the convergence of the algorithms to an invariant set of the ODE (6). We emphasize the asymptotics by including a superscript ε on the parameters when the gain is ε.

Theorem 3.1. *Under Assumptions A.1 in Appendix A.2, as $\varepsilon \to 0$, $\{\theta_t^\varepsilon\}$ converges weakly to a process that satisfies (6). Define the tail of the belief sequence by a shift q_ε with $\varepsilon q_\varepsilon \to \infty$ as $\varepsilon \to 0$. Then $\{\theta_t^\varepsilon\}_{q_\varepsilon}^\infty$ converges weakly to a limit set of (6).*

The convergence theorems show agents' beliefs converge to a limit set of the ODE (6), but there may be many such sets. In particular, when there are multiple stable equilibria which are limits of the learning rule, the convergence results do not distinguish among them. However since the same stability criteria apply in the discounted and un-discounted cases, we can use results from the literature which establish global convergence to a limit point in some special games. We call an equilibrium point $\bar{\theta} = (\overline{\theta_1}, \overline{\theta_2})$ of the ODE *linearly stable* if the real parts of all the eigenvalues of the Jacobian matrix of the belief dynamic ODE $\frac{\partial \bar{b}}{\partial \theta}(\bar{\theta})$ are strictly negative. A point is linearly unstable if at least one eigenvalue has strictly positive real part. We denote the set of linearly stable points by $\overline{\Theta}$. The following result summarizes known results from the literature.

Theorem 3.2. *The limit sets of the ODE (A.1) consist of points $\bar{\theta} \in \overline{\Theta}$ if:*

1. the game is 2×2,

2. the game is zero sum,

3. the game is a partnership game: the payoff matrices $(A = [a_{i,j}, B = b_{j,i}])$ satisfy: $x \cdot Ay = y \cdot Bx$ for all x and y, or

4. the game is cooperative: all off-diagonal elements of the Jacobian matrix of the belief dynamic ODE $\frac{\partial \bar{b}(\theta)}{\partial \theta}$ *are nonnegative.*
Therefore the tail $\{\theta_t^\varepsilon\}_{q_\varepsilon}^\infty$ *of the belief sequence converges weakly to a point in* $\overline{\Theta}$ *as* $\varepsilon \to 0$ *and* $\varepsilon q_\varepsilon \to \infty$*.*

Proof. Part 1 follows from Fudenberg and Kreps (1993) for a unique mixed equilibrium and Kaniovski and Young (1995) for the general case. (See also Benaim and Hirsch (1999a).) Parts 2 and 3 follow from Hofbauer and Hopkins (2000). Part 4 follows from Benaim and Hirsch (1999b) and Benaim (2000). The conclusions then follow from Theorem 3.1. ∎

In the special cases covered by Theorem 3.2, the learning algorithm will converge globally to a linearly stable point. In games that do not satisfy these conditions, there still is positive probability of convergence to a linearly stable point under some additional recurrence conditions, or convergence may at least be assured from appropriate initial conditions. A converse result (see Benaim and Hirsch (1999a)) also can be used to show that beliefs will not converge to an equilibrium point which is linearly unstable. In Section 4 below, we show how infrequent transitions between equilibria can lead to a particular stable equilibrium being the most likely outcome observed in the long run. But first we turn to two examples of stability analysis in the next section. The first illustrates the possibility of multiple stable equilibria, and the second examines the well-known Shapley (1964) game.

3.2. Stability Examples

3.2.1. A 2 × 2 Coordination Game

This example illustrates the possibility of multiple stable equilibria by examining a simple coordination game.

Example 3.1. Let the mean payoffs of a 2 × 2 coordination game be given by:

		Player 2	
		1	2
Player 1	1	3,3	2,0
	2	0,2	4,4

With normal shocks, the ODEs governing convergence are then:

$$\dot{\theta}_i = \Phi\left(\frac{5\theta_j - 2}{\sigma}\right) - \theta_i, \; i \neq j, \; i = 1, 2, \tag{7}$$

where Φ is the standard normal cumulative distribution function. Figure 1 shows the rest points for the ODE for this game (the logit case is very similar). The figure plots Φ from the ODE (7), so that equilibria are given by the intersections of Φ with the 45-degree line, and the stable equilibria are points where Φ intersects the line from above. For relatively large values of σ, the figure shows there is only one equilibrium,

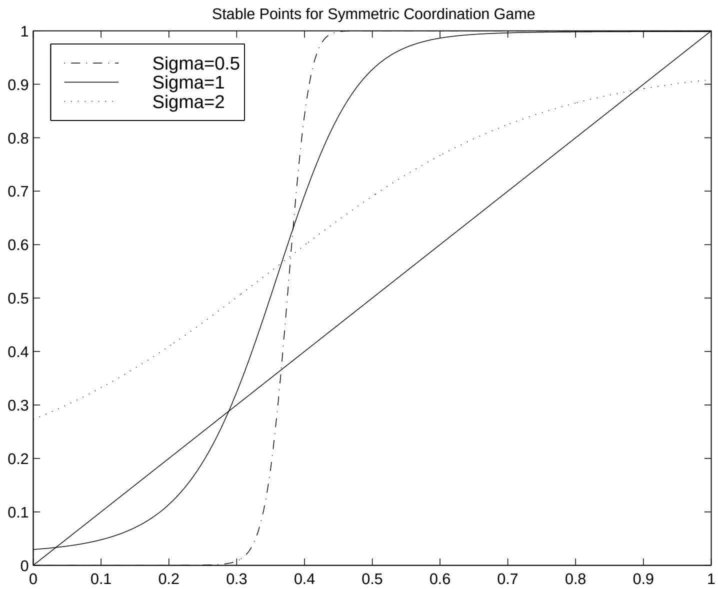

FIGURE 1. Stable rest points for a symmetric coordination game.

but as $\sigma \to 0$ there are three equilibria: two stable points in the neighborhood of the pure strategy Nash equilibria (at 0 and 1), and one unstable point in the neighborhood of the mixed strategy equilibrium (at 2/5 in this case). Thus Theorem 3.2 roughly implies that the tail of the belief sequence will converge to (a neighborhood of) one of the pure strategy equilibria, but it is silent on which of these outcomes to expect. We revisit this example below, where we show that one of the equilibria is more likely to be observed in the long run.

3.2.2. The Shapley (1964) Game

In this section, we use basic stability analysis of (A.1) in order to formally prove a conjecture of Fudenberg and Kreps's (1993) about the Shapley (1964) game. In discussing extensions of their results on 2×2 games, Fudenberg and Kreps (1993) stated that, "We suspect, however, that convergence cannot be guaranteed for general augmented games; we conjecture that an augmented version of Shapley's example will provide the desired counterexample, but we have not verified this." In this section we verify that the stochastic counterpart to Shapley's example is unstable if the shocks are small enough. Similar results for different games have been shown by Ellison and Fudenberg (2000) and Benaim and Hirsch (1999a).

Example 3.2. The payoffs in the Shapley game are given by the following:

		Player 2		
		1	2	3
Player 1	1	1,0	0,0	0,1
	2	0,1	1,0	0,0
	3	0,0	0,1	1,0

This game has a unique Nash equilibrium in which each player symmetrically plays all three actions with equal probabilities. In this deterministic game, Shapley (1964) proved that the fictitious play beliefs converge to a limit cycle in the three dimensional simplex. For the stochastic version of the game, we augment the payoffs by introducing the i.i.d. shocks as above. As we noted after Theorem 3.2, to show that an equilibrium is unstable it suffices to show that at least one the eigenvalues of $\frac{\partial B}{\partial \theta}$ has positive real part. The following theorem, with proof in the appendix, summarizes our results.

THEOREM 3.3. *In the normal case, for* $\sigma < 0.0821$ *and in the logit case for* $\lambda > 6$ *the unique mixed equilibrium* $\overline{\theta}_i = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, $i = 1, 2$ *in Example 3.2 is unstable.*

The calculations in the theorem also make clear that for large enough shocks, the equilibrium will be (at least) locally stable. While the result may not be surprising, it is interesting to note that the required noise in this example is not very large. For example, in the normal case the ratio of the mean payoff to the shock standard deviation is $0.333/0.0821 = 4.06$. Thus this is not a case of the noise simply swamping the mean payoff, as the "z-statistic" is highly significant. Thus it is possible that in games with stochastically perturbed payoffs players can learn to play equilibria, even when convergence fails in their deterministic counterparts.

4. ESCAPE AND STOCHASTIC STABILITY

In the previous section we provided conditions ensuring that agents' beliefs converge to a stable equilibrium. But we noted that convergence analysis alone can not determine which of many stable equilibria will be most likely to be observed. In this section we answer this question by characterizing the invariant distribution of beliefs in the constant gain case, showing that the limit distribution typically places point mass on a single *stochastically stable* (or long run) equilibrium. Asymptotically as the gain decreases (across sequences) agents' beliefs tend to spend most of their time within a small neighborhood of the stochastically equilibrium. We develop the results in this section at a general level, as they may be applicable to many different discrete time models of multiple equilibria and continuous state variables. This contrasts with the results in the literature which have focused on continuous time models (Foster and Young (1990), Fudenberg and Harris (1992)), or discrete state models (Kandori, Mailath, and Rob (1993), Young (1993)). As many economic models are naturally formulated in discrete time, our results have broad potential applications. In the next section we then apply our results to the stochastic fictitious play model.

On average, agents are drawn to a stable equilibrium, but occasional sequences of exogenous shocks may alter their assessments and cause them to change their strategies. This can cause the beliefs to "escape" from a stable equilibrium. These escape dynamics drive our characterization of stochastic stability. In our analysis below, we compute the probabilities that beliefs escape from one equilibrium to another, and therefore we determine a Markov chain over the set of stable equilibria. The invariant distribution of this chain determines the long run equilibrium.

Our characterization of stochastic stability relies on the discounted or constant gain nature of the learning rule. If, as in typical fictitious play specifications, the gain

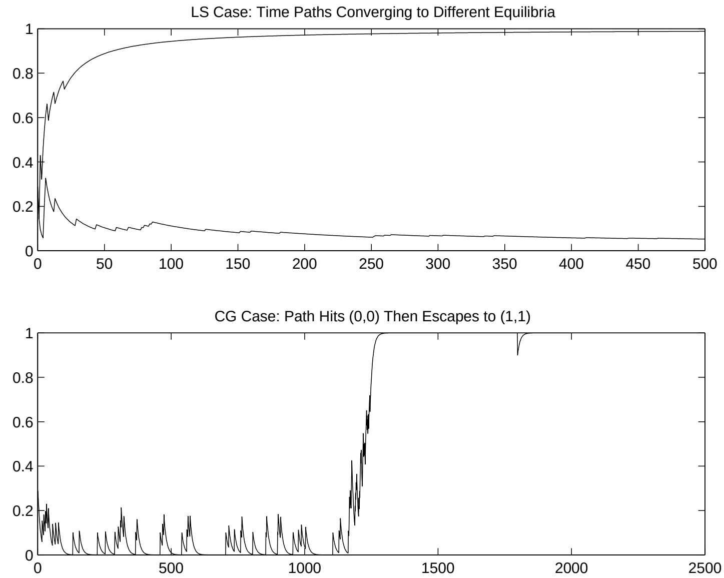

FIGURE 2. Simulated time paths from a coordination game under decreasing (LS) and constant gain (CG) learning.

decreased over time like $1/t$ (which we call least squares, or LS) then beliefs would tend to converge along a given sequence as time evolves. Thus the probability that the beliefs escape a stable equilibrium would go to zero along the sequence, so that there would not be sufficient "mixing" to determine a unique limit distribution. The outcome of the in that case would therefore be highly dependent on the initial condition and the particular realizations of shocks, and thus be difficult to characterize. However in with a constant gain case, we showed above that there is (weak) convergence in beliefs across sequences. This means that along any given trajectory of beliefs with a fixed gain setting, the probability that beliefs escape a given equilibrium remains nonzero. Thus we are able to deduce a stationary distribution that characterizes beliefs and is independent of the initial condition and specific shock realizations.

As an illustration of these issues, Figure 2 plots some simulated time paths from the coordination game in Example 3.1. Recall there are two linearly stable equilibria which are in neighborhoods of the symmetric pure strategy profiles. In the top panel, we plot two separate time paths of beliefs in the decreasing gain (LS) case, which start at the same initial value (which we set at the unstable mixed equilibrium) but converge to the two different stable equilibria. In the bottom panel, we plot a single time path in the constant gain (CG) case (again initialized at the unstable equilibrium), in which the beliefs are first drawn toward one of the equilibria and then escape to the other. The figure illustrates the difficulty in characterizing outcomes in the LS case and is suggestive of the possibility of characterizing stochastic stability in the CG case, which we pursue in this section.

In particular, we use techniques from large deviation theory to analyze the escape problem and calculate the stationary distribution. Our convergence theorems above suggest that the tail of beliefs tend to converge to a limit set. Therefore if we consider any event in which the beliefs start at a limit point and get arbitrarily far from it, that the event must have a probability converging to zero. However, as we have seen in Figure 2, for nonzero gain settings we do observe infrequent events where the beliefs move a substantial distance from a limit point. In this section we first establish a large deviation principle, due to Kushner and Yin (1997), which shows that the probability of these "escape" events decreases exponentially in the gain size and characterizes this rate of decrease. We then adapt results from Freidlin and Wentzell (1999) which use the large deviation principle to calculate an invariant distribution of beliefs. Finally we discuss how to implement the large deviation principle and calculate the stationary distribution, and therefore determine the stochastically stable equilibrium.

4.1. A Large Deviation Principle

In this section, we present a large deviation principle due to Kushner and Yin (1997), which draws on results by Dupuis and Kushner (1985, 1989). We present the theorem here using the general notation of Section 3.1 above, and in Section 5 below we show how to implement this general setup in the fictitious play model. In this section we assume that there is at least one point $\overline{\theta}$ which is a limit set for the ODE (6). The theorem provides upper and lower bounds on probability of observing an event in which agents' beliefs are arbitrarily far from a limit point. In particular, the theorem shows that once agents' beliefs are in the neighborhood of a stable equilibrium they remain there for an exponentially increasing period of time. In the next section we use this theorem to characterize the transition probabilities between equilibria, which then allows us to compute the invariant distribution of beliefs.

We now define some terminology and then develop some of the functions which are needed in statement of the theorem. Recall that $\overline{\Theta}$ is the set of stable limit points.

Definition 4.1. Fix an $\varepsilon > 0$, a time horizon $\bar{n} < \infty$ (which may depend on ε), and a compact set D with non-empty interior containing a stable limit point: $\bar{\theta} \in D$ for some $\bar{\theta} \in \overline{\Theta}$. Let $\theta^{\varepsilon}(t), t \in [0, \bar{n}\varepsilon]$ be the piecewise linear interpolation of $\{\theta^{\varepsilon}_t\}$.

1. An *escape path* from D is a sequence $\{\theta^{\varepsilon}_n\}_{n=0}^{\bar{n}}$ solving (3) such that $\theta^{\varepsilon}_0 = \bar{\theta}$ and $\theta^{\varepsilon}_m \notin D$ for some $m \le \bar{n}$. Let $\Gamma^{\varepsilon}(D, \bar{n})$ be the set of escape paths.

2. For any sequence $\{\theta^{\varepsilon}_n\}_{n=0}^{\bar{n}}$ solving (3) with $\theta_0 = \bar{\theta}$, define the (first) *escape time* from D as:

$$\tau^{\varepsilon}(\{\theta^{\varepsilon}_n\}) = \varepsilon \inf\{m : \theta^{\varepsilon}_m \notin D\} \in \mathbb{R} \cup \{\infty\}.$$

For small gains, any path $\{\theta^{\varepsilon}_n\}$ that starts near a stable limit will spend an increasing fraction of time near it, and if noise pushes it away, it tends to be drawn back. An escape path is a sequence of beliefs with exits a set D before the terminal date $\bar{n}$. The key results characterize bounds on the probability of escape and the the rate of increase of the escape times, which we use to characterize the distribution over stable equilibria.

In the theorem to follow, we characterize escapes by solving a cost minimization problem, and the functions that we develop now are the elements of that problem. First, the main function in our analysis is given by:

$$\begin{aligned} H(\theta,\alpha) &= \log E \exp \langle \alpha, b(\theta, e_t)\rangle \\ &= \left\langle \alpha, \bar{b}(\theta)\right\rangle + \log E \exp \langle \alpha, v_t\rangle, \end{aligned} \tag{8}$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product and the second equality uses equation (4). This function is simply the logarithm of the moment generating function of the $b(\theta, e_t)$ process, which is known as the cumulant generating function. Then we take the Legendre transform of H :

$$L(\theta, \beta) = \sup_{\alpha} \left[\langle \alpha, \beta \rangle - H(\theta, \alpha) \right]. \tag{9}$$

In the theorem that follows, L plays the role of an the instantaneous cost function, which "charges" paths that deviate from the stable point. With the flow cost L, we define a cumulative cost S for time paths $\mathbf{x} = (x(s))_0^T$:

$$S(T, \mathbf{x}) = \int_0^T L(x(s), \dot{x}(s)) ds.$$

The H function is the Hamiltonian from a cost minimization problem and the minimized cost function provides an estimate of the convergence rate. We spell out this Hamiltonian interpretation of H more fully in Section 4.3 below.

We now state our large deviation theorem, which is a known result from Kushner and Yin (1997). Let $\mathcal{B}_x$ be the set of continuous functions on a finite horizon $[0, T]$ taking values in the set D with initial condition x. Recall that for a set $\mathcal{B}_x$, $\mathcal{B}_x^0$ denotes is interior while $\overline{\mathcal{B}}_x$ denotes its closure. We analyze escapes on a fixed continuous time horizon $T < \infty$, and set $\bar{n} = T/\varepsilon$. Thus $\bar{n} \to \infty$ as $\varepsilon \to 0$. The large deviation principle is given in the following theorem. The necessary assumptions are collected in Assumptions A.2 in Appendix A, along with a brief discussion of a proof.

Theorem 4.1. *Assume that Assumptions A.2 in Appendix A hold, that the gain $\varepsilon > 0$ is constant, and that the shocks e_t are i.i.d. Let $\theta^\varepsilon(s)$ be the piecewise* linear *interpolation of $\{\theta_t^\varepsilon\}$. Then for a set D as in Definition 4.1 and for $T < \infty$:*

$$\begin{aligned} -\inf_{\mathbf{x} \in \mathcal{B}_x^0} S(T, \mathbf{x}) &\leq \liminf_{\varepsilon \to 0} \varepsilon \log P(\theta^\varepsilon(\cdot) \in \mathcal{B}_x \,\|\, \theta^\varepsilon(0) = x) \\ &\leq \limsup_{\varepsilon \to 0} \varepsilon \log P(\theta^\varepsilon(\cdot) \in \mathcal{B}_x \,\|\, \theta^\varepsilon(0) = x) \\ &\leq -\inf_{\mathbf{x} \in \overline{\mathcal{B}}_x} S(T, \mathbf{x}). \end{aligned} \tag{10}$$

The theorem shows that there is exponential decay (as $\varepsilon \to 0$) in the probability that beliefs will be far from a limit point. If the S function is continuous in the size of the escape set, then the limits in (10) exist and the inequalities become equalities. For example, if we define:

$$\overline{S} = \inf_{\left\{ \begin{array}{c} \mathbf{x} : x(0) = \overline{\theta}, \\ x(s) \notin D \text{ for some } s < T \end{array} \right\}} S(s, \mathbf{x}), \tag{11}$$

from (10) we have that:

$$P(\theta^\varepsilon(s) \notin D \text{ for some } 0 < s \leq T \,\|\, \theta^\varepsilon(0) = \overline{\theta}) = o\left(\exp(-\frac{1}{\varepsilon} \overline{S}) \right).$$

In addition, if we let τ^ε be the time of first escape from the set D, (10) implies that there exists some c_0 such that for small ε:

$$E\tau^\varepsilon \approx c_0 \exp(\frac{1}{\varepsilon}\overline{S}). \tag{12}$$

Thus the mean escape times increase exponentially in $\frac{1}{\varepsilon}$ as $\varepsilon \to 0$. In the next section we use (10) to calculate the transition probabilities between stable equilibria, which leads to a characterization of the asymptotic invariant distribution.

4.2. The Invariant Distribution and Stochastic Stability

In this subsection we adapt a theorem from Freidlin and Wentzell (1999), henceforth FW, to characterize the invariant distribution of beliefs. FW analyzed models of continuous time diffusions with small noise statistics. Although our development is in discrete time, the large deviation principle of the previous section allows us to extend the FW results to our present model. Kandori, Mailath, and Rob (1993) and Young (1993) similarly adapted the FW analysis for discrete-time, discrete state space models. (We focus on Young's work which is more general and closer to our analysis.) The model of this paper covers the intermediate case in which time is discrete but the state space is continuous, and thus it is not surprising that the FW/Young results extend to this case.

In the previous section we stated a large deviation principle for paths which start near a stable equilibrium and escape a set containing the equilibrium. In this section, we modify this analysis slightly to calculate the probability of transition between equilibria. Specifically, we now assume that there are K distinct stable equilibria that are the only limit sets of the ODE (18), as in the case of Theorem 3.2. Thus we have: $\overline{\Theta} = \left\{\overline{\theta}^1, ..., \overline{\theta}^K\right\}$. Then to deduce the asymptotic stationary distribution of the learning process, we can restrict our attention to this finite state space. Similar to the function $\overline{S}$ in (11) above, we define the following minimized cost function for paths connecting the stable equilibria $\overline{\theta}^i$ and $\overline{\theta}^j$:

$$V_{ij} = \inf_{\left\{\begin{array}{c} \mathbf{x} : x(0) = \overline{\theta}^i, \\ x(T) = \overline{\theta}^j \text{ for some } T < \infty, \\ x(s) \neq \overline{\theta}^k,\ k \neq i, j \text{ for } s \in (0, T) \end{array}\right\}} S(T, \mathbf{x}). \tag{13}$$

Above we showed that the asymptotic probability of escape was determined by the function $\overline{S}$, and FW show (Lemma 6.2.1) that the asymptotic transition probabilities of the Markov chain on $\overline{\Theta}$ are determined by V. Thus the invariant distribution of beliefs can be computed from these transition probabilities, as we now establish. First, following FW (p.177) and analogous to Young (1993), we define the following.

Definition 4.2. For a subset $\omega \subset \overline{\Theta}$, a graph consisting of arrows $\overline{\theta}^m \to \overline{\theta}^n$ $\left(\overline{\theta}^m \in \overline{\Theta} \setminus \omega, \overline{\theta}^n \in \overline{\Theta}, n \neq m\right)$ is called a ω-**graph** if it satisfies the following:

(i) every point $\overline{\theta}^m \in \overline{\Theta} \setminus \omega$ is the initial point of exactly one arrow,
(ii) there are no closed cycles in the graph.

We denote the set of ω-graphs by $\mathcal{G}(\omega)$, and define the following function:

$$W_i = \min_{g \in \mathcal{G}\left(\overline{\theta}^i\right)} \sum_{\left(\overline{\theta}^m \to \overline{\theta}^n\right) \in g} V_{mn}. \tag{14}$$

In words, this function looks at the sum of the costs V of transitions between equilibria along all graphs anchored at $\overline{\theta}^i$ and chooses the minimizing graph. When there are only two stable equilibria, there is a single graph for each equilibrium and thus $W_i = V_{ji}$. In cases with more stable equilibria, the number of graphs proliferates and the calculation of W becomes slightly more complex. (See Young (1993) for an example.)

Finally, let μ^ε be the invariant measure of the beliefs $\{\theta_n^\varepsilon\}$. Then we have the following theorem which identifies the long run equilibrium. It follows directly from Theorem 6.4.1 of FW, with our Theorem 4.1 replacing FW Theorem 5.3.2. All of the supporting lemmas required in the proof of the theorem can be seen to hold in the current model.

Theorem 4.2. *Assume that the conditions of Theorem 4.1 above hold, and that limit sets of the ODE A.1 are a finite set $\overline{\Theta}$ of stable equilibria. Then for any $\gamma > 0$ there exists a $\rho > 0$ such that the μ^ε-measure of the ρ-neighborhood of the equilibrium $\overline{\theta}^i$ is between:*

$$\exp\left(-\varepsilon\left(W_i - \min_j W_j \pm \gamma\right)\right)$$

for sufficiently small ε.

The theorem implies that as $\varepsilon \to 0$ the invariant measure of beliefs μ^ε is concentrated in a small neighborhood of the equilibria that attain the minimum of W. If the minimum is attained at a unique equilibrium, then the invariant distribution asymptotically places all of its mass within a small neighborhood of this equilibrium. These results justify the following definition.

Definition 4.3. A **stochastically stable equilibrium** is a stable equilibrium $\overline{\theta}$ that satisfies $\overline{\theta} \in \arg\min_j W_j$ and so is in the support of the invariant distribution μ^ε as $\varepsilon \to 0$.

Thus, a stochastically stable (or long run) equilibrium is an outcome that is likely to be observed in the long run evolution of the system. In particular, if there is a unique stochastically stable equilibrium, then as the gain ε decreases we expect to observe the agents' beliefs spending an increasing fraction of time within a small neighborhood of it. Starting from an arbitrary initial point, by Theorem 3.1 we know that the beliefs are drawn to one of the stable equilibria. However as time passes, by Theorem 4.1 we know that there is a nonzero probability that the beliefs will eventually escape this equilibrium and be drawn to another. As this process continues through time, Theorem 4.2 establishes that the beliefs eventually spend most of the time near the stochastically stable equilibrium. It is in this sense that stochastic stability provides a selection criterion for models with multiple stable equilibria. In order to characterize the stochastically stable equilibrium, we then need to determine the escape paths that lead from one equilibrium to another. In Theorem 4.1, and in the definition of V in (13) above, we saw that the escape paths solve control problems, and in the next section we further characterize the solutions to these control problems.

4.3. Characterizing the Long Run Equilibrium

Following Fleming and Soner (1993), we can find an analytical expression for the differential equations that characterize the dominant escape paths between equilibria, which requires one further assumption.

Assumption 4.2. *L is strictly convex in β and obeys a superlinear growth condition*:

$$\frac{L(\theta,\beta)}{|\beta|} \to \infty \text{ as } |\beta| \to \infty.$$

Under this assumption, we can characterize the solution of the calculus of variations problem for the escape problem as the solution to appropriate differential equations. This is just an application of Pontryagin's maximum principle, with the resulting first order conditions and adjoint equations. Notice that H and L are convex duals, so that similar to (9) we have:

$$H(\theta,\alpha) = \sup_{\beta}\left[\langle \alpha,\beta\rangle - L(\theta,\beta)\right].$$

The cost minimization problem which characterizes an escape path between equilibria and determines the value V_{ij} can then be written:

$$V^j(x) = \inf_{\mathbf{x}} \int_0^T L(x(s), \dot{x}(s))ds \tag{15}$$

$$\text{s.t. } x(0) = x,\ x(T) = \overline{\theta}^j \text{ for some } T < \infty. \tag{16}$$

Then $V_{ij} = V^j(\overline{\theta}^i)$. The Hamiltonian for this problem with state x, co-state λ, and control $\dot{x}$ is:

$$-\mathcal{H}(x,\lambda) = \inf_{\dot{x}}\left\{L(x,\dot{x}) + \lambda\cdot \dot{x}\right\} = -H(x,a), \tag{17}$$

where $a = -\lambda$. Thus we see that the Hamiltonian is the H function that we defined above. Further, by taking the appropriate derivatives of the Hamiltonian we see that the dominant escape path can be found as the solution to the differential equations:

$$\begin{aligned} \dot{x}(s) &= H_\alpha(x(s), a(s)) \\ \dot{a}(s) &= -H_\theta(x(s), a(s)), \end{aligned} \tag{18}$$

subject to the boundary conditions (16).

Alternatively, following Fleming and Soner (1993) we can characterize the solution of the cost-minimization problem (15)-(16) by dynamic programming methods.[6] If we

[6] See also Dembo and Zeitouni (1998), Exercise 5.7.36.

let $V_x^j(x) = \partial V^j(x)/\partial x$, we have that the value function $V^j(x)$ satisfies the following Hamilton-Jacobi partial differential equation:

$$H(x, -V_x^j(x)) = 0, \tag{19}$$

where we've used (17).

We can solve the key cost minimization problem (15)-(16) using either the PDE characterization in (19) or the ODE characterization in (18). In some special cases, the PDE characterization leads to explicit analytic results, as we show below. However for general multidimensional games, we must rely on numerical solutions, and the ODE characterization is easier to implement numerically. To see this, note that we can solve (18) as a two-point boundary problem with the given initial condition for x and a terminal condition $x(T_0) = \overline{\theta}^j$ for some T_0. This solution gives *an* escape path from $\overline{\theta}^i$ to $\overline{\theta}^j$. In other words, adapting notation, by solving the boundary value problem we obtain a path that allows us to calculate:

$$V_{ij}(T_0) = \inf_{\left\{\mathbf{x}:\, x(0)=\overline{\theta}^i,\, x(T_0)=\overline{\theta}^j\right\}} \int_0^{T_0} L(x(s), \dot{x}\ (s))ds.$$

In order to find the most likely escape path, we minimize this function over $T < \infty$:

$$V_{ij} = \inf_{T<\infty} V_{ij}(T). \tag{20}$$

Thus we have determined the asymptotic transition probabilities between equilibria. By using these values to calculate the W function as in (14), we can then determine the invariant distribution and find the stochastically stable equilibrium. In the special case where there are two stable equilibria, we have already seen that the W function reduces to V. We collect these results in the following corollary.

Corollary 4.1. *Suppose there are two stable equilibria: $\overline{\Theta} = \left\{\overline{\theta}^1, \overline{\theta}^2\right\}$. Then $\overline{\theta}^1$ is the stochastically stable equilibrium if $V_{12} > V_{21}$, where $V_{ij} = \int_0^T L(x(s), \dot{x}\ (s))ds$, L is defined in (8) and (9), $x(t)$ solves (18) subject to (16), and T is the minimizer in (20).*

5. STOCHASTIC STABILITY IN FICTITIOUS PLAY

In this section we apply the results of the previous section to the model of stochastic fictitious play. First we derive some of the key functions in the 2×2 case and indicate how to extend the results to higher dimensions. Under stochastic fictitious play, the cost function L simplifies and has a natural interpretation. Then we establish that in 2×2 symmetric games the stochastically stable equilibrium is the risk dominant equilibrium. Finally, we turn to some illustrations of our results via numerical calculations and simulations. We first illustrate the main theorem of this section on the coordination game from Example 3.1. Here we see that it is extremely difficult to switch from the risk dominant equilibrium to the other. Then we turn to a 3×3 coordination game due to Young (1993), where our results differ from that paper. The persistent stochastic nature of our model causes one of the equilibria to become unstable, which changes the nature of stochastic stability.

5.1. Application to Fictitious Play

Here we specialize the results of the previous section by explicitly deriving the key functions in the fictitious play model. We first focus on the 2×2 case with two stable equilibria and then discuss extensions to larger dimensions. Due to the discrete nature of the learning algorithm in (2), the H function is rather simple to compute. Let $\theta = (\theta_1, \theta_2)'$ and as in Appendix A.1, define $G^1(\theta_2, i)$ as the probability that player 1 plays action i with similar notation for player 2. As we only keep track of a single probability for each agent, we use the notation $G^1(\theta_2) = G^1(\theta_2, 1)$ and $G^2(\theta_1) = G^2(\theta_2, 1)$. By the independence of the $e^i_{j,t}$ we have that:

$$
\begin{aligned}
H(\theta, \alpha) &= \log E \exp\left(\alpha_1 \left(1_{\{\text{Player 1 plays 1}\}} - \theta_1\right)\right) + \\
&\quad \log E \exp\left(\alpha_2 \left(1_{\{\text{Player 2 plays 1}\}} - \theta_2\right)\right) \\
&= -\left(\alpha_1\theta_1 + \alpha_2\theta_2\right) + \log\left[1 + \left(\exp\left(\alpha_1\right) - 1\right) G^1(\theta_2)\right] \\
&\quad + \log\left[1 + \left(\exp\left(\alpha_2\right) - 1\right) G^2(\theta_1)\right].
\end{aligned}
\tag{21}
$$

Next, recall that the relative entropy of the discrete distribution p with respect to the discrete distribution q is defined as:

$$
I(p, q) = \sum_{i=1}^{N} p_i \log \frac{p_i}{q_i}.
$$

Further, recall that I is nonnegative and equal to zero only when $p = q$. Then using (9) and (21) we see that L takes the form:

$$
L(\theta, \beta) = I(\beta_1 + \theta_1, G^1(\theta_2)) + I(\beta_2 + \theta_2, G^2(\theta_1)). \tag{22}
$$

Recall that the "cumulative cost" S of a potential path from one equilibrium to another is measured by the integral of $L(\theta, \dot{\theta})$ along the path. The properties of the entropy function ensure that the instantaneous cost is zero only along paths which follow the ODEs (18) governing convergence. Therefore to escape an equilibrium requires the beliefs to overcome the force pushing them back toward the equilibrium, which entails a cost. Further, the most likely escape paths are those paths between stable equilibria that minimize the cumulative relative entropy between what an agent believes and what his opponent believes about him.

To find the most likely escape path, we then solve the control problem (15) for each of the two equilibrium transitions. Corollary 4.1 implies the stochastically stable equilibrium is the equilibrium $\overline{\theta}^i$ with the larger value of V_{ij}. By Theorem 4.1 above, the value of $-V_{ij}$ provides an estimate of the log probability that the beliefs will escape from $\overline{\theta}^i$ to $\overline{\theta}^j$, and so the stochastically stable equilibrium is the equilibrium with the lower escape probability.

In the next section we use the PDE characterization in (19) to deduce that in symmetric 2×2 games the stochastically stable equilibrium is the risk dominant equilibrium. However to obtain quantitative results on the speed of equilibrium transitions, we can solve the problem using the ODE characterization in (18). The differential equations which determine the evolution of the states and co-states along the escape path are then given by:

$$\begin{aligned} \dot{\theta}_i &= H_{\alpha_i}(\theta, \alpha) = -\theta_i + \frac{\exp(\alpha_i) G^i(\theta_j)}{1 + (\exp(\alpha_i) - 1) G^i(\theta_j)} \\ \dot{\alpha}_j &= -H_{\theta_j}(\theta, \alpha) = \alpha_j - \frac{(\exp(\alpha_i) - 1) \frac{\partial G^i(\theta_j)}{\partial \theta_j}}{1 + (\exp(\alpha_i) - 1) G^i(\theta_j)}, \end{aligned} \tag{23}$$

for $i, j = 1, 2,\ i \neq j$. We solve these differential equations subject to the boundary conditions (16).

In order to characterize stochastic stability in larger games, we simply extend this analysis. In these cases the PDE generally characterization becomes more difficult, so we use the ODE characterization to obtain numerical results. The calculation of the H function is straightforward, and simply involves replacing $G^i(\theta)$ by its multidimensional counterpart in an extension of (21). The differential equations that characterize the escape paths are also the obvious corollaries of (23), with the complications that the derivatives of the G^i are more difficult to evaluate, and that the H_θ derivatives have additional cross-effect terms.

5.2. Stochastic Stability in Symmetric 2 × 2 Games

In this section we analyze the important special case of symmetric 2×2 games. Here we show that the stochastically stable equilibrium is the risk dominant equilibrium. As we discuss below, this result is the same as several other evolutionary selection criteria in the literature. However in Section 5.4, we show that this equivalence does not necessarily extend beyond the 2×2 case.

We now suppose that the payoff matrices A and B are identical, which clearly implies that player's beliefs are driven by identical dynamics. Therefore, since the equilibria are symmetric, we can focus on a single state variable $\theta = \theta_1 = \theta_2$, driven by the function $G(\theta) = G^1(\theta) = G^2(\theta)$. We consider the only relevant case where there are three (perturbed) equilibria, two stable $(\overline{\theta}^1, \overline{\theta}^2)$ and one unstable $(\widetilde{\theta})$. Without loss of generality, we suppose that the equilibria can be ordered as:

$$0 \leq \overline{\theta}^2 < \widetilde{\theta} < \frac{1}{2} < \overline{\theta}^1 \leq 1. \tag{24}$$

Therefore $\overline{\theta}^1$ is the risk-dominant equilibrium and correspondingly has the larger basin of attraction. From (A.1) we see that if we initialize the mean dynamics at $\theta < \widetilde{\theta}$, we get $\theta(t) \to \overline{\theta}^2$ and vice versa.

The main theorem of this section is based on the analysis of the cost minimization problem (15) and its PDE characterization (19). In the special case of this section, by using (21), the PDE (19) determining the value functions V^1 and V^2 reduces to a system of differential equations:

$$-V_x^j(x)x = \log\left[1 + \left(\exp\left(-V_x^j(x)\right) - 1\right) G(x)\right], \quad j = 1, 2. \tag{25}$$

Even in this relatively simple setting, it is typically not possible to determine an analytic solution for this equation (25). However we can use it to establish properties of the two solutions. The proof of the following theorem relies on showing that derivative of $V^1(x)$ is uniformly greater than $V^2(x)$ over the relevant regions defined by (24). When coupled with the larger basin of attraction of $\overline{\theta}^1$, by Corollary 4.1 this implies that $\overline{\theta}^1$ is the stochastically stable equilibrium. The proof is given in Appendix A.5.

THEOREM 5.1. *In symmetric* 2×2 *games, the stochastically stable equilibrium is the risk dominant equilibrium.*

Our results in this section agree with the equilibrium selection criteria in Young (1993) and Kandori, Mailath, and Rob (1993). In fact our analysis agrees with all of the selection criteria discussed by Kim (1996). The different criteria have different dynamic adjustment processes, but in each case the risk dominant equilibrium has the largest basin of attraction. In the discrete state models of Young and KMR, mutations are of a fixed size. Therefore a larger basin means that more needed mutations to escape to the equilibrium, which means that escape is less likely. In our setting, payoff perturbations are of varying size, so there is a trade-off between receiving a few large shocks and an accumulation of small shocks. Thus although we cannot simply count mutations, the same intuition results. To escape from the risk dominant equilibrium requires a "more unlikely" string of perturbations. Kim (1996) showed that the equivalence between different selection criteria in the 2×2 case does not extend to multi-player games, and we show below that for larger two player games the criteria differ. But first we give a quantitative illustration of our results.

5.3. A 2×2 Example: A Coordination Game

As an example, we return to the symmetric coordination game from Example 3.1 above. As above, we focus on the cases when the shocks are normally distributed with standard deviation σ or have extreme value distribution with parameter λ. The three rest points and escape probabilities for the two stable equilibria with different settings of σ and λ are given in Table 1 below. The limit points in the table are rounded to 4 digits, and are not exactly zero or one as the perturbations ensure they remain fully mixed.

From the table, we first see that (as in Figure 1) as the variance of the shocks decreases, the rest points become closer to the Nash equilibria. We also see that, in accord with Theorem 5.1, the log escape probability is much lower at $\overline{\theta}^2$ than $\overline{\theta}^1$, and thus $\overline{\theta}^2$ is the stochastically stable equilibrium. This means that if agents start out coordinating on action 1, which is the Pareto-inferior equilibrium, then the probability that they will switch to the Pareto optimal equilibrium decays rapidly to zero as the gain decreases. In contrast, although the probability of switching from the Pareto superior equilibrium to the inferior one also declines to zero, for any given nonzero gain setting there is a much greater chance of switching in this direction. Thus we expect that, in the long run, the equilibrium in which both players play action 1 with

TABLE 1.

Long run equilibrium in a 2×2 coordination game.

Model	Rest Points			$-$Log Prob.	
	$\overline{\theta}^1_{i1}$	$\widetilde{\theta}_{i1}$	$\overline{\theta}^2_{i1}$	V_{12}	V_{21}
$\sigma = 1.0$	0.033	0.288	0.998	0.279	3.712
$\sigma = 0.7$	0.002	0.344	1	1.432	6.699
$\sigma = 0.5$	0	0.366	1	3.168	11.681
$\lambda = 5$	0	0.381	1	2.023	5.729
$\lambda = 10$	0	0.391	1	4.515	11.663
$\lambda = 20$	0	0.396	1	9.280	22.296

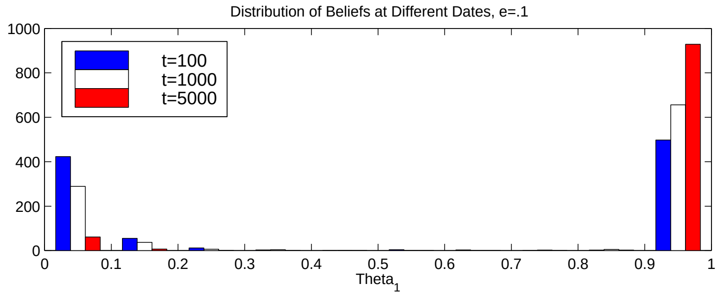

FIGURE 3. Long run evolution of beliefs in a 2×2 coordination game.

probability near one will be observed much more frequently than the other stable equilibrium.

We illustrate the evolution of beliefs in Figure 3, which shows a histogram of beliefs from 1000 samples from the normal case with $\sigma = 1$. Each sample was initialized at the unstable equilibrium and was run for 5000 time periods. Just as our analysis suggests, the beliefs are initially drawn to a neighborhood of one of the pure strategy equilibria. After 100 periods, the samples are nearly equally divided between $\overline{\theta}^1$ and $\overline{\theta}^2$. However as time goes by, the beliefs eventually escape $\overline{\theta}^1$ and are drawn to $\overline{\theta}^2$, which they have a very low probability of escaping. By the time we reach period 5000 nearly the entire mass of the distribution is concentrated near $\overline{\theta}^2$. Thus in the long run evolution of the game, we see that the beliefs eventually converge to the stochastically stable equilibrium and stay there for an increasingly long period of time.

5.4. Long Run Equilibrium in a 3×3 Game

We next present an example that illustrates a difference between our model and Young's (1993) selection criteria. In this example, Young's criterion selects a Pareto inferior (and non-risk-dominant) equilibrium in a coordination game, while our criterion selects the Pareto optimal (and risk dominant) equilibrium. Young's criterion supports its choice by transitions through a third coordination equilibrium that yields the lowest payoffs, but in our model this equilibrium is very seldom visited. In fact, when the payoff shocks are large enough (but still relatively small in magnitude), this worst equilibrium disappears. For small enough shocks, the equilibrium emerges but has a very small basin of attraction. We find that it is difficult to escape to this equilibrium and very easy to escape from it. The difference between the two models, as we discuss further below, deals with the nature of the perturbations. Rather than "mutations" causing agents to randomly choose an action, agents in our model choose the best action based on the realizations of their perturbed payoffs. Thus in order to play an action with a low expected payoff, an agent must observe a sequence of large shocks which make that action look more favorable. The dependence of the perturbed choice probabilities on the payoffs drives the difference in the results.

Our analysis in this section is necessarily numerical. Unlike our analysis of stability above, we cannot provide precise analytic characterizations of the equilibria and their stability. In those examples, the expected payoffs of all strategies were symmetric, which greatly simplified the analysis. In this section, we will examine games whose strategies have different mean payoffs. We can still determine analytically the set of equilibria in the limit as the noise goes to zero, but for any strictly positive shock variance we must determine the perturbed equilibria and their stability properties numerically.[7] Furthermore, unlike the previous example, the game that we analyze here does not fall into the subset of games covered by Theorem 3.2. Thus we cannot be assured that the limit sets of the ODE consist only of stable points. However numerical analysis of the differential equation suggests that this is so (see Figure 4 below), and so we proceed under this assumption. We now turn to computing the long run equilibrium in an example coordination game due to Young (1993).

Example 5.1. Let the expected payoffs in a 3×3 coordination game be given by:

		Player 2		
		1	2	3
	1	6,6	0,5	0,0
Player 1	2	5,0	7,7	5,5
	3	0,0	5,5	8,8

In the augmented form of this game, it is easy to show that in the limit as the noise gets small, all of the coordination equilibria are stable. This holds for both the normal and logit cases. Therefore for small enough shocks, the set of stable equilibria are near the pure strategy coordination equilibria:

$$\overline{\Theta} = \left\{ \overline{\theta}^1, \overline{\theta}^2, \overline{\theta}^3 \right\}, \text{ where } \overline{\theta}_i^1 \approx (1,0,0), \ \overline{\theta}_i^2 \approx (0,1,0), \ \overline{\theta}_i^3 \approx (0,0,1).$$

However numerical results show that for a range of strictly positive shock variances there are only two stable equilibria, $\overline{\theta}^2$ and $\overline{\theta}^3$. This is particularly the relevant in the normal case. Recall that for an equilibrium, $\bar{b}(\theta) = 0$ so we take $\left\|\bar{b}(\theta)\right\|$ as a measure of the distance an arbitrary strategy is from being an equilibrium. For example, for $\sigma = 0.3$ we have $\left\|\bar{b}((1,0,0))\right\| = 0.013$, $\left\|\bar{b}((0,1,0))\right\| = 1.2 \times 10^{-6}$, and $\left\|\bar{b}((0,0,1))\right\| = 7.68 \times 10^{-13}$. The values for the pure equilibria 2 and 3 are very small, and we were able to find perturbed equilibria near them. However we could find no such perturbed equilibrium $\overline{\theta}^1$ for this noise specification.

[7] This amounts to calculating the integrals in $B\left(\overline{\theta}\right)$ and the eigenvalues of $\frac{\partial B}{\partial \theta}(\overline{\theta})$, which are simple numeric calculations.

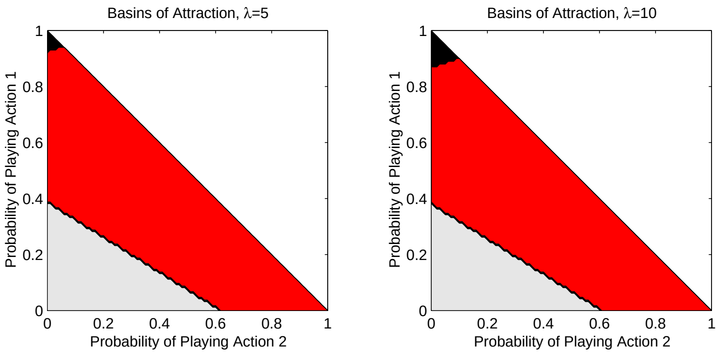

FIGURE 4. The basins of attraction of the equilibria in Example 5.1 in the logit case for different specifications of the shocks.

For the logit case, for the parameters we considered we were able to find three perturbed equilibria close to the pure equilibria. As an illustration, Figure 4 plots the basins of attraction for the different equilibria in the logit case for different values of λ. In the figure the equilibria are located at the corners of the triangle. The basins of attraction are the shaded areas around the equilibria, which we determined by solving the ODE (18) numerically. Here we see that for the larger value of the noise ($\lambda = 5$), the basin of attraction of $\overline{\theta}^1$ (the shaded area in the upper left of the figure) is very small and that it increases slightly as the noise decreases. To preview the results to follow, the figure suggests that it should be relatively easy to escape from $\overline{\theta}^1$, and therefore that the results rely mainly on comparing $\overline{\theta}^2$ and $\overline{\theta}^3$. We see below that we obtain similar results in both the normal and logit cases, indicating that this worst coordination equilibrium plays a relatively small role in our analysis.

We then solve the cost minimization problems as described above in order to determine the stochastically stable equilibrium. In particular, we consider the normal case in which there are only two stable equilibria ($\overline{\theta}^2$ and $\overline{\theta}^3$), and the logit case in which all three equilibria are stable. Table 2 summarizes our results. The table lists the noise specifications and the log transition probabilities between each of the equilibria. For the normal cases, since the stochastic nature of the payoffs rules out one of the possible equilibria, we can apply Corollary 4.1. Thus the stochastically stable equilibrium is the one with the higher value of V and so the lower escape probability. The table shows that the stochastically stable equilibrium in this example is the Pareto dominant equilibrium $\overline{\theta}^3$. For the logit cases, we must find the minimum cost graph as in Theorem 4.2 above. From the table, it is easy to see that the minimizing graph is anchored at $\overline{\theta}^3$, and involves the arrows $\overline{\theta}^1 \to \overline{\theta}^2 \to \overline{\theta}^3$ for a total cost of $W_3 = V_{12} + V_{23}$. Thus again we find that the stochastically stable equilibrium is the risk dominant and Pareto optimal equilibrium $\overline{\theta}^3$.

Similar to the 2×2 case, we then verify these results by tracking the evolution of beliefs over time in some simulations. In Figure 5 we plot the distribution of beliefs from 1000 simulated time paths at different dates in the normal case with $\sigma = 0.5$. Here we use the gain setting $\varepsilon = 0.3$, and initialize beliefs randomly on the simplex. In the top panel we plot the second element $\theta_{1,2}$ of the belief vector, and the bottom panel plots the distribution of third element $\theta_{1,3}$. We see that initially the beliefs

TABLE 2.

Equilibrium transition rates in a 3×3 coordination game.

Model	$-$Log Escape Prob.					
	V_{12}	V_{13}	V_{21}	V_{23}	V_{31}	V_{32}
$\sigma = 0.5$	0	-	-	2.4	-	7.4
$\sigma = 0.3$	0	-	-	5.0	-	15.8
$\lambda = 5$	0.14	3.4	27.0	2.3	24.7	5.9
$\lambda = 10$	0.78	7.1	52.8	4.6	48.3	11.7

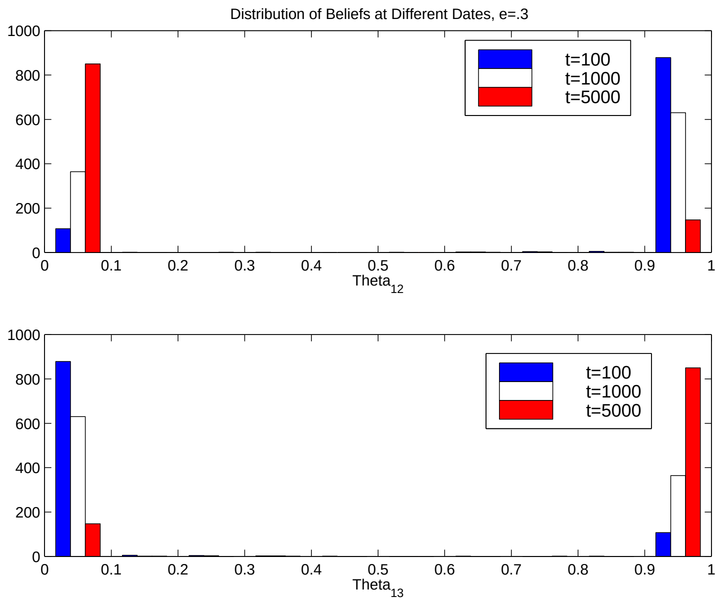

FIGURE 5. Long run evolution of beliefs in a 3×3 coordination game.

are drawn to one of the stable equilibria, which in our simulations initially favors $\overline{\theta}^2$. However over time the beliefs eventually escape this equilibrium and are drawn to the stochastically stable equilibrium $\overline{\theta}^3$.

As we've noted, in this example our selection criterion differs from Young's and agrees with Pareto and risk dominance criteria. It is easy to check that in this example, the equilibrium $\overline{\theta}^1$ risk dominates the other two coordination equilibria in pairwise comparisons, and therefore it is the risk dominant equilibrium. However, Young's criterion selects the equilibrium $\overline{\theta}^2$, which relies on transitions through the worst equilibrium $\overline{\theta}^1$. For the normal cases we considered, the stochastic shocks ensure that agents never converge to this coordination equilibrium with the lowest payoffs. By ruling out this equilibrium, we are then essentially left with a 2×2 pure coordination game, and as described above our criterion selects a risk dominant equilibrium in such a game. The modifier *essentially* is important in this statement because we do not restrict the players to only play the two strategies. However we rule out one of the equilibria as a possible limit point, and so calculate the transitions between the two pure strategy equilibria. In our analysis, the third strategy was never chosen along these transitions. For any nonzero gain, the third strategy may be chosen along the path of play, but the frequency it is played converges to zero.

This intuition also carries over to the logit cases, where all three of the equilibria are stable. As we saw in Figure 4, the basin of attraction of the worst coordination equilibrium $\overline{\theta}^1$ is very small. Table 2 shows that is very easy to escape this equilibrium and converge to $\overline{\theta}^2$, as V_{12} is very low. Thus the minimum graph problem essentially reduces to comparing V_{23} and V_{32} as in the normal case.

As discussed above, the differences between our model and the model of Young (1993) rely on the nature of the perturbations. In Young's paper, a perturbation or mistake leads agents to choose a random action, with each alternative being equally likely. Thus in his analysis of this game, the easiest way to transit from $\overline{\theta}^3$ to $\overline{\theta}^2$ is to have an opponent choose action 1 by mistake a sufficient number of times. In our model, the perturbations hit agents' payoffs, and agents select the best option after observing the realizations of shocks. Starting at $\overline{\theta}^3$, an agent assesses very low expected payoffs to action 1 relative to action 2. Therefore extremely large shocks are necessary for the agent to choose this action. Much more likely is the case that action 2 becomes viewed more preferably. Thus in our model, the transitions between equilibria are direct. The choice probabilities in our model directly reflect the payoffs, instead of there being a chance that agents choose a completely random action. This directed choice has important consequences for equilibrium selection.

6. CONCLUSION

In this paper we have presented a general method for analyzing models with multiple stable equilibria, and have applied it to the stochastic fictitious play model of learning in games. Our methods focus on individual agents, who myopically optimize based on their beliefs. By introducing stochastic shocks to their payoffs, and assuming that agents discount past observations when they learn, we derived criteria for long run equilibrium selection. In particular, we showed that the stochastic nature of our model along with the directed choice drive our results. Sufficiently large stochastic shocks may rule out cycling that prevents convergence in some games, and thus can lead to stability of mixed equilibria. In addition, the stochastic shocks rule out certain equilibria which drive some of the results in the evolutionary literature. Further, the fact that agents choose actions to maximize perturbed payoffs, instead of occasionally choosing completely random actions, means that very large shocks are needed to play actions with very low payoffs. This directly affects the probabilities that equilibria are played in the long run. Therefore although our methods are related to existing results in the learning and evolutionary game theory literature, they are based on different principles and lead to some different results.

The main analytic results in this paper were developed at a general level, and our methods have broad potential applications for economic models with multiple equilibria. Many such models are naturally cast in the discrete-time continuous state stochastic framework and can now be analyzed using our methods. The fundamental point of our results is that in models with a multiplicity of equilibria, not all equilibria are equal. If agents must learn the structure of the economy, then limit points of their learning rules will be most likely observed. Even among those equilibria which are stable under learning, an equilibrium which is the easiest to learn and which takes the longest time to escape will be the most likely outcome observed in the long run.

APPENDIX A
Assumptions and Proofs

A.1. CONSTRUCTING THE DIFFERENTIAL EQUATION

Here we derive the differential equation which govern the convergence of beliefs. To find the limit ODE, we take expectations in (2), which means that we must calculate the probabilities that each player plays each action. In the 2×2 case, this is a simple calculation leading to a probit decision rule in the normal case and a logit rule in the extreme value case. However in larger dimensions, the calculations become more complex. For example, to compute the probability that player 1 chooses action i, note that the term on the right-hand side of (1) is the $(N-1)$ order statistic from an independent but not identically distributed sample of size $(N-1)$. We denote this order statistic as $X(\theta_{2,t}, i)$, to emphasize the dependence on the agent's beliefs $(\theta_{2,t})$, and the reference strategy (i in this case). The order statistics for the two players have the following cumulative distribution functions (see David, 1970):

$$\begin{aligned} F^1_{X(\theta_{2,t},i)}(x) &= \prod_{j\neq i}^{N} F\left(x - \theta_{2,t} \cdot a_j\right) \\ F^2_{X(\theta_{1,t},i)}(x) &= \prod_{j\neq i}^{N} F\left(x - \theta_{1,t} \cdot b_j\right). \end{aligned}$$

Therefore the probabilities that player 1 plays action i and player 2 plays action j are respectively given by:

$$\begin{aligned} G^1(\theta_{2t}, i) &= \int_{-\infty}^{\infty} F^1_{X(\theta_{2t},i)}(x) dF\left(x - \theta_{2,t} \cdot a_i\right) \\ G^2(\theta_{1t}, j) &= \int_{-\infty}^{\infty} F^2_{X(\theta_{1t},j)}(x) dF\left(x - \theta_{1,t} \cdot b_j\right) \end{aligned}$$

These expressions are rather complicated, and do not lead to explicit evaluation when the shocks are normally distributed. However when the shocks have the extreme value distribution, it is well known that the probability that player 1 plays action i takes the form:

$$G^1(\theta_{2,t}, i) = \frac{\exp(\lambda\theta_{2,t} \cdot a_i)}{\sum_{j=1}^{N} \exp(\lambda\theta_{2,t} \cdot a_j)}.$$

In the discrete choice econometrics literature, this is known as a multinomial logit model. Such a specification was used by McKelvey and Palfrey (1995) for their notion of a quantal response equilibrium. Fudenberg and Levine (1995) also derived an identical choice rule based on deterministic perturbations of payoffs.

With these calculations, the ODEs governing convergence can be written explicitly:

$$\begin{aligned} \dot{\theta}_{1i} &= G^1(\theta_2, i) - \theta_{1,i} \\ \dot{\theta}_{2j} &= G^2(\theta_1, j) - \theta_{2,j}, \end{aligned} \tag{A.1}$$

for $i, j = 1, ..., N$. Since agents' beliefs are constrained to lie on the unit simplex, we can reduce the dimensionality of the state space. Therefore, we define $\bar{b}(\theta)$ from (18) as the $(2(N-1) \times 1)$ vector composed of the right side of (A.1) for the first $N-1$ elements of θ_1 and θ_2.

A.2. ASSUMPTIONS FOR CONVERGENCE

We first briefly describe the continuous time approximation. Recall that convergence is as $\varepsilon \to 0$ across sequences indexed by the gain setting. To distinguish between discrete and continuous time scales, we now let n be the discrete time index. For the continuous time scale, let $t_0 = 0$ and $t_n = n\varepsilon$. Let $\{q_\varepsilon\}$ be a sequence of nonnegative, nondecreasing integers, and define for $t \geq 0$:

$$Z^{\varepsilon,q}(t) = \varepsilon \sum_{i=q_\varepsilon}^{t/\varepsilon + q_\varepsilon - 1} b(\theta_i^\varepsilon, e_i),$$

where the integer part of t/ε is used in the limit of the summation. Then $\theta^\varepsilon(\varepsilon q_\varepsilon + t) = \theta^\varepsilon_{q_\varepsilon} + Z^{\varepsilon,q}(t)$ is the right-continuous, piecewise-constant, continuous time shifted process associated with $\{\theta^\varepsilon_n\}$.

The following conditions lead to the weak convergence result of Kushner and Yin (1997), Theorem 8.5.1. Since we only consider i.i.d. shocks some of their additional assumptions are immediate.

Assumptions A.1.

(i) $\{\theta^\varepsilon_n; \varepsilon, n\}$ *is tight.*[1]

(ii) For each compact set Q, $\left\{b(\theta^\varepsilon_n, e_n)1_{\{\theta^\varepsilon_n \in Q\}}; \varepsilon, n\right\}$ *is uniformly integrable.*[2]

(iii) For each compact set Q, *the sequence* $\left\{\bar{b}(\theta^\varepsilon_n)1_{\{\theta^\varepsilon_n \in Q\}}; \varepsilon, n\right\}$ *is uniformly integrable.*

(iv) There are nonempty compact sets S_i *that are the closures of their interiors* S_i^0 *and satisfy* $S_0 \subset S_1^0 \subset S_1 \subset S_2^0 \subset S_2$ *such that all trajectories of the ODE (6) tend to* S_0 *as* $t \to \infty$ *and all trajectories starting in* S_1 *stay in* S_2. *Further, the ODE (6) has a unique solution for each initial condition.*

(v) The function $\bar{b}(\theta)$ *is continuous.*

(vi) For each $\delta > 0$, *there is a compact set* A_δ *such that* $\inf_{n,\varepsilon} P(\upsilon^\varepsilon_n \in A_\delta) \geq 1 - \delta$.

The result then follows from Kushner and Yin (1997), Theorem 8.5.1, where the assumptions are easily verified by noting that the belief sequence is bounded and that the G^i functions are continuous.

A.3. STABILITY EXAMPLE

Proof (Theorem 3.3). It is easy to verify that $\bar{\theta}$ is an equilibrium point of the ODE, and we now verify that it is linearly unstable. After much calculation, we see that the Jacobian at $\bar{\theta}$ takes on the following form:

$$\frac{\partial \bar{b}}{\partial \theta}(\bar{\theta}) = \begin{bmatrix} -1 & 0 & \overline{G} & 0 \\ 0 & -1 & 0 & \overline{G} \\ 0 & \overline{G} & -1 & 0 \\ -\overline{G} & -\overline{G} & 0 & -1 \end{bmatrix}$$

where $\overline{G}$ depends on the shock distribution. Further, the four eigenvalues δ of $\frac{\partial \bar{b}}{\partial \theta}(\bar{\theta})$ solve:

$$(-1-\delta)^2 = \pm \overline{G}^2.$$

As long as $\overline{G} > 2$ all of the eigenvalues have strictly positive real parts, which is clearly sufficient for instability.

In the logit case, $\overline{G} = \frac{\lambda}{3}$, so we require $\lambda > 6$ for instability. In the normal case, we define $z = \frac{x - 1/3}{\sigma}$ and:

$$\begin{aligned} \overline{G} &= \int_{-\infty}^{+\infty} \Phi(z)\phi^2(z)dx + \int_{-\infty}^{+\infty} \Phi^2(z)\phi(z)zdx \qquad \text{(A.2)} \\ &= \left(1 + \frac{1}{2\sigma}\right)\int_{-\infty}^{+\infty} \Phi^2(z)\phi(z)zdx \\ &\equiv \left(1 + \frac{1}{2\sigma}\right)\widehat{G}, \end{aligned}$$

where $\widehat{G} = 0.2821$ is a constant independent of σ. Thus we require $\sigma < \frac{\widehat{G}}{2(2-\widehat{G})}$. Evaluating the right side numerically gives the result. ▌

[1]A random sequence $\{A_n\}$ is *tight* if:

$$\lim_{K\to\infty} \sup_n P(|A_n| \geq K) = 0.$$

[2]A random sequence $\{A_n\}$ is *uniformly integrable* if:

$$\lim_{K\to\infty} \sup_n E\left(|A_n| 1_{\{|A_n| \geq K\}}\right) = 0.$$

A.4. LARGE DEVIATIONS

The following assumptions from KY are necessary for Theorem 4.1 above.

ASSUMPTIONS A.2.

KY, A6.10.1. The ODE (6) has a unique solution for each initial condition, and a point $\overline{\theta}$ that is locally (in D) asymptotically stable. The function $b(\cdot, e_n)$ is bounded and right-continuous in θ.

KY, A6.10.4. The real-valued function $H(\theta, \alpha)$ in (8) is continuous in (θ, α) for $\theta \in D$ and α in a neighborhood Q of the origin. The $\alpha-$derivative of H is continuous on Q for each fixed $\theta \in \overline{D}$.

Under these assumptions, Theorem 4.1 is an extension of a result in KY, as the following sketch makes clear.

Proof (Theorem 4.1). Follows from KY, Theorem 6.10.2, with the discussion on pp.177-178 or the derivation in Dupuis and Kushner (1989) identifying the H function. The stated result in KY also requires that b be Lipschitz continuous, but the analysis in Dupuis and Kushner (1985) shows that right-continuity and boundedness are sufficient. ▮

A.5. PROOF OF THE **2** × **2** THEOREM

Proof (Theorem 5.1). We first note a few elementary facts about the value functions $V^j(x)$ from the minimization problem (15). Note that since L is increasing and convex, $V^j(x)$ is also increasing and convex (in the distance from $\overline{\theta}^j$). Further, it is clear that $V^j(\widetilde{\theta}) = 0$ since the mean dynamics initialized on either side of $\widetilde{\theta}$ converge to the closest equilibrium. Also note that even though the equilibria and dynamics are symmetric, we could allow for asymmetric escape paths. However the convexity of L implies that such paths would not be minimizing.

The proof uses the PDE characterization from (25), which here is simply an ODE. To conserve slightly on notation we consider the case when $\overline{\theta}^1 = 1$ and $\overline{\theta}^2 = 0$. This clearly holds for small noise, and the more general case only involves a change in notation. To compare the solutions, we re-orient the state space in each case. That is, for V^1 we use the transformation $x = \widetilde{\theta} - y$ and for V^2 we transform as $x = \widetilde{\theta} + y$. Therefore at $y = 0$ both start at $x = \widetilde{\theta}$ and both are increasing in y. Then, using our results above, we can write solutions as the following:

$$\begin{aligned} V_{12} &= V^2(\overline{\theta}^1) = \int_0^{1-\widetilde{\theta}} V_y^2(y)dy, \\ V_{21} &= V^1(\overline{\theta}^2) = \int_0^{\widetilde{\theta}} V_y^1(y)dy, \end{aligned}$$

where the derivatives with respect to the transformed variables y are easily deduced from the original and the transformations.

From our results above, we have that $V_y^j \geq 0$ for $j = 1, 2$. Then since V_{12} involves integrating over a larger area, if we can show $V_y^2(y) \geq V_y^1(y)$ for $y \in (0, \widetilde{\theta})$ we then have $V_{12} > V_{21}$. This is what we now establish. First, we use the transformations to re-write the two ODEs from (25) as:

$$\begin{aligned} a(y)(\widetilde{\theta} + y) &= \log\left[1 + (\exp(a(y)) - 1)\, G(\widetilde{\theta} + y)\right] \\ b(y)(\widetilde{\theta} - y) &= \log\left[1 + (\exp(b(y)) - 1)\, G(\widetilde{\theta} - y)\right] \end{aligned} \tag{A.3}$$

where $a(y) = -V_y^2(y)$ and $b(y) = -V_y^1(y)$, so $a, b \leq 0$. Note that (A.3) implicitly defines a, b for a given y, and further that $a(0) = b(0) = 0$ and $a(1 - \widetilde{\theta}) = 0$ while $b(\widetilde{\theta}) = 0$. That is, $a(y)$ and $b(y)$ are non-positive, are zero at $\widetilde{\theta}$, and $b(y)$ hits zero before $a(y)$. Therefore, if $a(y)$ and $b(y)$ do not intersect on $(0, \widetilde{\theta})$ we must have $a(y) \leq b(y)$ on that interval, which would complete the proof.

To establish this via a contradiction, suppose that $a(y) = b(y)$ for some $y \in (0, \widetilde{\theta})$. Then by (A.3) we have:

$$1 + (\exp(a(y)) - 1)G(\widetilde{\theta} + y) = \exp\left(\frac{\widetilde{\theta} + y}{\widetilde{\theta} - y}\right)\left[1 + (\exp(a(y)) - 1)G(\widetilde{\theta} - y)\right],$$

which in turn implies:

$$\exp(a(y)) - 1 = \frac{1 - \exp\left(\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y}\right)}{G(\widetilde{\theta}+y) - \exp\left(\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y}\right) G(\widetilde{\theta}-y)}. \tag{A.4}$$

To insure that $a(y)$ is non-positive and real, we require $-1 \leq \exp(a(y)) - 1 \leq 0$. Since we have:

$$\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y} \geq 1 \text{ for } y \in (0, \widetilde{\theta}),$$

by (A.4) we then require:

$$\begin{aligned} G(\widetilde{\theta}+y) - \exp\left(\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y}\right) G(\widetilde{\theta}-y) &\geq 0, \text{ and} \\ G(\widetilde{\theta}+y) - \exp\left(\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y}\right) G(\widetilde{\theta}-y) &\leq 1 - \exp\left(\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y}\right). \end{aligned}$$

The second inequality can be re-written as:

$$\frac{1 + G(\widetilde{\theta}+y)}{1 + G(\widetilde{\theta}-y)} \geq \exp\left(\frac{\widetilde{\theta}+y}{\widetilde{\theta}-y}\right). \tag{A.5}$$

Now, on $(0, \widetilde{\theta})$ the left side of (A.5) takes on a value of at most 2, while the right side takes on values of at least $\exp(1) > 2$. Thus there is no value of y on the interval so that (A.5) holds. Therefore $a(y)$ and $b(y)$ do not intersect on the interval, which completes the proof. ▌

REFERENCES

1. Benaim, Michel (1999) "Dynamics of Stochastic Approximation Algorithms, " *Sminaire de Probabilits, XXXIII*, Lecture Notes in Mathematics 1709, Springer, 1-68.
2. Benaim, Michel (2000) "Convergence with Probability One of Stochastic Approximation Algorithms whose Average is Cooperative," *Nonlinearity,* 13, 601-616.
3. Benaim, Michel and Morris W. Hirsch (1999a) "Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games," *Games and Economic Behavior,* 29, 36-72.
4. Benaim, Michel and Morris W. Hirsch (1999b) "Stochastic Approximation Algorithms with Constant Step Size whose Average is Cooperative," *Annals of Applied Probability,* 9, 215-241.
5. Benaim, Michel and Jorgen Weibull (2003) "Deterministic Approximation of Stochastic Evolution in Games", *Econometrica,* 71, 873-903.
6. Binmore, Ken, Larry Samuelson and Richard Vaughn (1995) "Musical Chairs: Modeling Noisy Evolution," *Games and Economic Behavior* 11, 1-35.
7. Boylan, Richard T. (1995) "Continuous Approximation of Dynamical Systems with Randomly Matched Individuals," *Journal of Economic Theory* 66, 615-625.
8. Bravo, Mario (2016) "An Adjusted Payoff-Based Procedure for Normal Form Games," *Mathematics of Operations Research*, 41: 1469-1483.
9. Brown, G.W. (1951) "Iterative Solutions of Games by Fictitious Play," in *Activity Analysis of Production and Allocation* (T.C. Koopmans, ed.), New York: John Wiley & Sons.
10. Cheung, Yin-Wong and Daniel Friedman (1997) "Individual Learning in Normal Form Games," *Games and Economic Behavior,* 19: 46-76.
11. Cominetti, Roberto, Emerson Melo, and Sylvain Sorin (2010) "A Payoff-Based Learning Procedure and its Application to Traffic Games," *Games and Economic Behavior*, 70: 71-83.
12. David, H.A. (1970) *Order Statistics,* New York: John Wiley & Sons.
13. Dembo, Amir and Ofer Zeitouni (1998) *Large Deviations Techniques and Applications,* Second Edition, Springer-Verlag, New York.
14. Dupuis, Paul and Harold J. Kushner (1985) "Stochastic Approximation via Large Deviations: Asymptotic Properties," *SIAM Journal on Control and Optimization,* 23, 675-696.

15. Dupuis, Paul and Harold J. Kushner (1989) "Stochastic Approximation and Large Deviations: Upper Bounds and w.p.1. Convergence," *SIAM Journal on Control and Optimization,* 27, 1108-1135.
16. Ellison, Glenn and Drew Fudenberg (2000) "Learning Purified Mixed Equilibria," *Journal of Economic Theory,* 90, 84-115.
17. Fleming, Wendell and H. Mete Soner (1993) *Controlled Markov Processes and Viscosity Solutions,* Springer-Verlag, New York.
18. Foster, Dean and H. Peyton Young (1990) "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology,* 38, 219-232.
19. Freidlin, M.I. and A.D. Wentzell (1999) *Random Perturbations of Dynamical Systems,* Second Edition, Springer-Verlag, New York.
20. Fudenberg, Drew and Christopher Harris (1992) "Evolutionary Dynamics with Aggregate Shocks," *Journal of Economic Theory,* 57, 420-441.
21. Fudenberg, Drew and David M. Kreps (1993) "Learning Mixed Equilibria," *Games and Economic Behavior,* 5, 320-367.
22. Fudenberg, Drew and David K. Levine (1995) "Consistency and Cautions Fictitious Play," *Journal of Economic Dynamics and Control,* 19, 1065-1089.
23. Harsanyi, John (1973) "Games with Randomly Disturbed Payoffs," *International Journal of Game Theory,* 2, 1-23.
24. Hofbauer, Josef and Ed Hopkins (2000) "Learning in Perturbed Asymmetric Games," manuscript, University of Edinburgh.
25. Hofbauer, Josef and William Sandholm (2002) "On the Global Convergence of Stochastic Fictitious Play," *Econometrica,* 70, 2265-2294.
26. Kaniovski, Yuri M. and H. Peyton Young (1995) "Learning Dynamics in Games with Stochastic Perturbations," *Games and Economic Behavior,* 11, 330-363.
27. Kandori, Michihiro, George Mailath and Rafael Rob (1993) "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica,* 61, 29-56.
28. Kim, Youngse (1996) "Equilibrium Selection in n-Person Coordination Games," *Games and Economic Behavior*, 15, 203-227.
29. Kushner, Harold J. (1984) "Robustness and Approximation of Escape Times and Large Deviations Estimates for Systems with Small Noise Effects," *SIAM Journal on Applied Mathematics,* 44: 160-182.
30. Kushner, Harold J. and George G. Yin (1997) *Stochastic Approximation Algorithms and Applications,* Springer-Verlag, New York.
31. Leslie, David S. and E.J. Collins (2006) "Generalised Weakened Fictitious Play," *Games and Economic Behavior*, 56: 285298.
32. McKelvey, Richard D. and Thomas R. Palfrey (1995) "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior,* 10, 6-38.
33. Myatt, David P. and Chris Wallace (2004) "Adaptive Play by Idiosyncratic Agents," *Games and Economic Behavior*, 48: 124-138.
34. Robinson, J. (1951) "An Iterative Method of Solving a Game," *Annals of Mathematics,* 54, 296-301.
35. Sandholm, William H. and Mathias Staudigl (2016) "Large Deviations and Stochastic Stability in the Small Noise Double Limit," *Theoretical Economics,* 11, 279-355.
36. Sarin, Rajiv and Farshid Vahid (1999) "Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice," *Games and Economic Behavior,* 28, 294-309.
37. Shapley, L.S. (1964) "Some Topics in Two-Person Games," in *Advances in Game Theory* (M. Dresher, L.S. Shapley, and A.W. Tucker, eds.), Princeton University Press.
38. Williams, Noah (2018) "Escape Dynamics in Learning Models," *Review of Economic Studies*, forthcoming.
39. Young, H. Peyton (1993) "The Evolution of Conventions," *Econometrica*, 61, 57-84.