Average Behavior in Learning Models

DAVID CANNING

*Cambridge University, United Kingdom and
Columbia University, New York, New York 10027*

We examine a general class of adaptive behavior models in which the distant past has only a weak effect on current actions, and assume that agents sometimes make mistakes, to show that average behavior (averaged over time) converges, with probability one, to a unique limit. Mistakes generate global convergence and are an equilibrium selection device; for small mistake probabilities the equilibrium selected is close to an equilibrium of the model without mistakes. The overlapping generations model, and learning in games with bounded memory, fit into this framework and are examined as examples of the result. *Journal of Economic Literature* Classification Numbers: D80, C62. © 1992 Academic Press, Inc.

1. INTRODUCTION

The assumption that agents have perfect information, though frequently used in economic theory, is often not satisfied in practice. While useful as a bench-mark, to discover which effects do not depend on imperfect information, realism dictates that we go further and investigate models with imperfect information.

One approach to the problem of imperfect information is to assume that agents know the structure of the "true model" but that some relevant parameters of the model are random and unobserved. This view is based on Savage's [23] idea that the agent knows all the possible "states of the world" but is uncertain about which have, or will, occur. In addition, we often assume that the agent knows the objective probability distribution over states of the world and, given this prior belief, uses Bayes' rule to update whenever he observes some new information. Harsanyi's [17] Bayesian–Nash equilibrium uses this method to deal with imperfect information in game theory.

There are three drawbacks to this approach; we must assume that agents know all the possible states of the world, that they know the objective probability distribution of the random draws made over these states, and that they have the computational capability to solve the rather complex problem of applying Bayes' rule to such a large state space. To highlight these problems we need only consider the likely size of the set of states of

442

the world. In game theory it must consist of all the possible information sets that could be given to each agent; we may be uncertain not only about the structure of the model but also about what our opponents' know. Part of what our opponents' have information about is our knowledge; we must form beliefs over their beliefs about our beliefs. They, in turn, will have beliefs over these beliefs of ours, and so on, given an infinite hierarchy of beliefs. While this problem can be resolved, using common knowledge assumptions (see Brandenburger and Dekel [2]), it is difficult to imagine agents going very far down this route in practice.

A more intuitively appealing approach is to assume that agents follow behavioral rules that are less complex than full Bayesian learning. A behavioral rule should tell the agent what actions to take, and how to change these actions in the light of experience. The behavioral rule we use in economics, expected utility maximization ombined with Bayesian updating, requires too much of the agents. At the other extreme biologists (see Maynard Smith [20]) take the animal's choice of action as fixed and instinctive, changes in behavior occurring only at the level of the population due to the higher reproductive rates of animals using successful strategies. In economics we like to think of human behavior as being a little more sophisticated than the instinctive choices of animals. People do change their behavior with new experience and we do observe adaptive behavior, though perhaps not of the kind predicted by full Bayesian learning.

It is natural to assume that agents learn by adapting to the variables they observe. In the Savage approach agents form beliefs about the fundamentals of the problem and their subjective probabilities on particular events occurring are given by the probability they put on the fundamentals being such that the event happens. If we wish to guess if someone will press a button or not the Savage approach is to form probabilities over all possible "states of mind" of that person and add together the probabilities on all those states which lead him to press the button. An alternative approach is to consider probabilities as frequencies, and to assign as the probability that this agent will press the button the frequency with which we have observed people press the button in the past. This approach is applied to learning in games by Canning [6] and Fudenberg and Levine [15]. We can think of this type of learning as "superficial"; agents have beliefs over what will happen; they do not care, or think much about, why things happen. This concentration on observables, and probabilities (or frequencies) on observables, certainly makes calculation easier for the agents; however, it does give rise to the problem that the agent's model of how events are linked may be misspecified.

The fundamental problem with learning using such an approach is the non-stationarity of the observable variables. If the observed variables are

stationary then a simple learning rule based on this stationarity will, eventually, learn the distribution and subjective beliefs will be close to the underlying objective probability distribution. However, in situations in which many agents interact, and learn, non-stationarity arises naturally from the fact that agents are observing each other, and each is changing his behavior as he learns. Bray and Savin [3] investigate a linear rational expectations model in which agents learn by using ordinary least squares and they show that convergence to a steady state depends on the parameters of the model.

In this paper, instead of focusing on a particular model, we attempt to derive conditions under which general learning models will give convergence. The two key behavioral assumptions we require are that the agents actions are based on boundary memory (or that the distant past does not matter "too much") and that they sometimes make mistakes, choosing an action that is independent of their history. These conditions imply that the empirical distribution of outcomes in the model converges to an unique limit distribution. If we average outcomes over time this empirical average converges, with probability one, to the average of the limit distribution, which is independent of the initial conditions of the model. Additionally, we can consider a model made up of large number of "villages" or "islands," agents interacting within each village, but with independence between villages. Averaging over villages, at sufficiently large times, will give average behavior close to the average of the limit distribution.

In the proofs of my results I rely heavily on the convergence theorems for stochastic dynamical systems given by Futia [16]. An alternative approach is to follow Norman [21] and study the system as a random contraction (a contraction "on the average"). This has the advantage of giving a clearer intuition for the results; different histories given rise to different future behavior, but mistakes, being independent of history, tend, when they occur, to bring different paths for the system "closer together." I the long run, the possibility of mistakes, even if they are unlikely, tends to lead to convergence and this force dominates over the divergent paths the system might otherwise follow. While more intuitive, and allowing more direct proofs, the random contraction approach has only been developed for the case where the set of possible outcomes is bounded. In many interesting economic models the set of possible outcomes is unbounded (e.g. the explosive price paths in speculative bubbles) and it seems desirable to include this possibility in the analysis. The random contraction approach is used in this paper in the case where outcomes are bounded when it allows us to relax the bounded memory assumption to allow actions which depend upon, but which are relatively insensitive to, the distant past.

The idea that stochastic dynamical systems have better convergence properties than deterministic systems appears in Young [26], Evans [10], and Canning [5]. These papers all use a similar approach; the equilibrium of the system is an absorbing state, once such an absorbing state occurs the system is stuck there forever. Out of equilibrium the system moves randomly, but has a positive probability of entering an absorbing state within some finite number of periods; in the long run it must enter such a state (the probability of being in an absorbing state approaches one as time passes). The model presented here differs in that with the possibility of mistakes there are no absorbing states; the system keeps being perturbed. Instead of converging to an absorbing state the system converges to a limiting, non-degenerate, probability distribution over states. Equilibrium, if thought of as a fixed point of the dynamic system, must be thought of not as a particular state but as a probability distribution over states.

In general the equilibrium behavior we derive depends on the probabilities with which agents make mistakes and what they do when they make a mistake. If the model was very sensitive to these factors we would not be able to use the results from a particular model with any confidence about their robustness. We therefore investigate the nature of the equilibrium distributions of behavior when the probability of mistakes is small. We show that for small probabilities of mistakes equilibrium behavior is close to an equilibrium distribution for the model without mistakes. Without mistakes we do not have existence, uniqueness, or convergence results. If we assume the set of possible outcomes is compact we can show existence of an equilibrium and, for small probabilities of mistakes, average behavior converges to a limit which is close to one of these equilibria. Mistakes, therefore, give global convergence and provide an equilibrium selection criterion. In particular, if, in the case of adaptive behavior without mistakes, we can show the existence of a unique equilibrium, mistakes will make this globally stable. This equilibrium selection criterion is intuitively similar to using randomly perturbed continuous time dynamical systems to select robust equilibria (see Freidlin and Wenzel [13]), a device applied by Foster and Young [12] in the context of biological evolution with mutation.

Adding a small probability of mistakes selects a unique equilibrium but which equilibrium is selected depends, in general, on the distribution of mistakes assumed. However, Young [27] examines a large class of learning models for which the equilibrium selected is independent of the mistake distribution providing it full support.

The power of our results is shown by applying them to some simple models. To begin with we investigate the game theory models of learning developed by Canning [6] and Fudenberg and Levine [15]. In a simple example of a repeated many player game, with two pure strategy equilibria,

we actually work out what the equilibrium distribution of behavior looks like and show that the risk dominant equilibrium (Harsanyi and Selten [18]) is stable. This result is similar to that found in Kandori, Mailath, and Rob [19]. Finally, we give an example showing how our assumptions that agents have finite memories and make mistakes can be replaced by the assumption that the model consists of overlapping generations; death limits memory and the newborn, having no information, tend to make "mistakes."

2. Average Behavior in Adaptive Models

We consider a society with n agents. At every time t each agent i has a set of possible actions A_i. The n vector of actions $a \in A = A_1 \times A_2 \times \cdots \times A_n$, together with a random component, determines an outcome $s \in S$. In an adaptive behavior model we assume that the action a_i an agent chooses at time t is a function of the history of outcomes up to that time.

This formulation may appear restrictive; however, as we shall see in our examples, it is not. The generality comes from the fact that we have not restricted the action or event spaces. For example, we can admit mixed strategy choices as "actions." Alternatively, in models where we may wish agents' actions to depend on current state variables, as well as history, we can take the "action" to be the submission of a function which says what to do depending on the current state. If agents observe only part of the relevant history, and there is incomplete or asymmetric information, we can take each agent's actions as a function which is constant over histories which lie in the same information partition.

To put some structure on the model we make the following assumptions.

Assumption 1.   Each A_i and S are Borel subsets of complete separable metric spaces.

This is a technical condition which will usually be satisfied in practice. For example, it is satisfied if the action and outcome spaces are finite dimensional Euclidean spaces. Agents actions depend on the history they have experienced. We restrict the choice of action by assuming either of the following.

Assumption 2a. $H = S^k$, for k finite, and $a_{it} = f_i(h_t)$ where $h_t = (s_{t-1}, s_{t-2}, ..., s_{t-k})$ and each $f_i: H \to A_i$ $(i = 1, 2, 3, ..., n)$ is continuous in the product topology on H.

or

Assumption 2b. S is compact, $H = S^\infty$, and $a_{it} = f_i(h_t)$ where $h_t =$

$(s_{t-1}, s_{t-2}, ...)$ and each $f_i : H \to A_i$ $(i = 1, 2, 3, ..., n)$ is continuous in the product topology on H.

The product topology on $H = S^\infty$ can be generated by the metric

$$d(h_t, h_t') = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{d_s(s_{t-i}, s_{t-i}')}{1 + d_s(s_{t-i}, s_{t-i}')},$$

where d_s is the metric on the space S. Under Assumption 2b we allow unbounded memory but note that the factor $1/2^i$ in the metric implies that outcomes in the distant past have little effect on current behavior.

Both Assumptions 2a and 2b require actions to be continuous in the agent's history. The difference lies in the fact that in 2a we require bounded memory while in 2b we have unbounded memory but require the outcome space to be compact.

The key to Assumption 2a is that there is a limit on how far back agents can remember. There is of course, in practice, the problem that while agents may have bounded lives, and hence bounded memory, information may be transmitted between generations by written, or other, messages. Indeed, an agent who knows he forgets things will often write messages to himself (e.g., in a diary or in files). The model is directly applicable only in circumstances, such as overlapping generation models where, for some reason, the new entrants either do not have, or do not take account of, evidence on the previous, distant, history. In practice, moral hazard problems may well give to circumstances where agents give little weight to what other agents tell them has happened, preferring to rely on personal experience.

If we include writing messages in the action space, and let the outcome space include receiving these messages, there is no loss of generality in asuming agents have bounded memory. We can embed the history in the current state of the system, making the current state space very large, but reducing the memory requirement. In practice when we "remember" the past we really only access a current state stored in our mind. Similarly books and historical artifacts, indeed any relics of the past, are current states. Sabourian [22] shows how this embedding can be carried out in repeated games with one period memory, without restricting the agents' abilities to implement strategies that normally depend on longer time spans. While such embedding appears to overcome the bounded memory assumption it is an illusory achievement once we add the possibility of mistakes; mistakes in repeatedly recording history to keep it in the current state eventually corrupt ancient history to the point where it is worthless.

We can drop the assumption of bounded memory providing we take the outcome space, S, to be compact and assume continuity of the actions in

a topology (such as the infinite product topology) over the infinite history which makes histories "close" if they are the similar in latest k components for some bounded k. This implies that agents take similar actions for histories which are similar in the recent past; the distant past does not affect current actions much.

The assumption that actions are continuous in histories appears restrictive. For example, in matrix games actions are not continuous in beliefs over the opponents' mixed strategy choice. In addition, if agents have information sets which are derived from a finite partition of a Euclidean history space then their actions cannot be continuous over histories unless they are independent of the history. However, these problems can be overcome if we assume that agents' observations of the history are noisy; that is, what they see is close to the actual outcome but randomly perturbed. This "trembling eye" generates a probability distribution over actions and it is easy to show that this probability distribution varies continuously with the history, and so satisfies our assumption. In this paper we do not use this technical device and simply take Assumption 2 as given. However, it is worth noting that the continuity condition always holds when the state space is countable; in the discrete topology no two histories can be close and so any function is continuous.

Taking a_t to be the vector of planned actions at time t we can write $a_t = f(h_t)$ where f is the vector of functions f_i and h_t is the history. Clearly, since each of the functions f_i is continuous f is continuous, under either 2a or 2b.

So far we have talked only about planned actions, actual actions may differ from planned actions due to a "trembling hand"; agents may make mistakes when choosing their strategy.

Assumption 3. With probability $(1 - p) < 1$ an agent plays his planned action. With probability p he plays randomly, choosing from the fixed distribution α_i on A_i where α_i has finite support.

I have assumed that the error probability is the same for each agent; this is not essential. All our results hold with different error probabilities for each agent so long as all these probabilities are strictly positive. The assumption that the mistake distributions have a finite support is also made for simplicity; our results hold for more general probability distributions, but this requires more complex proofs than given here.

The actions played by the agents depend on whether or not agents make mistakes. Given a planned vector of actions $f(h)$ (depending on the history h) we can calculate the probability measure $\theta(f(h))$ on A that results from the different possible combinations of mistakes. $\theta(f(h))$ will have an atom of weight $(1 - p)^n$ on the planned actions $f(h)$ and a weight of p^n on the measure generated by all the agents making a mistake. We also have to

take all the possible combinations among the n players of planned actions and mistakes to calculate $\theta(h)$ completely. The probability of an action in the set B given the history h is given by $(\theta(f(h)))(B)$; to avoid confusion we write this simply as $\theta_h(B)$ where θ_h denotes the probability measure generated by the probability p, the distribution α_i, and function f, which operates on the history h.

The idea that agents make mistakes can be justified, as in Selten [24], on straightforward "trembling hand" grounds; agents genuinely do make mistakes. Alternatively, we might think of the "mistakes" as part of a behavioral strategy; agents might be unsure of the payoffs to different actions and may experiment in an effort to determine these payoffs (see Fudenberg and Kreps [14]). Another possible justification of Assumption 3 is to consider an overlapping generations model, as in Fudenberg and Levine [15]; the newborn, with no information about history, must take an action which is history independent. In Section 6 we show how our assumption that agents make mistakes can be replaced with the assumption of overlapping generations.

Given an action at time t the outcome may depend on some random elements. We can regard S as a measurable space by taking S^* to be the σ-algebra generated by the Borel subsets of S. Let

$$g: A \to \Omega(S)$$

denote the mapping from the set of actions to the set of probability measures over the possible outcomes. If the model is deterministic then $g(a_t)$ will put an atom of weight one on the outcome; if the model is stochastic $g(a_t)$ will be a probability distribution over outcomes.

In what follows we use two notions of distance in the space of probability measures. Let w be a probability measure on the set X. First, the total variation norm on X is defined by (see Stokey and Lucas [25])

$$|w| = \sup_B \sum_{i=1}^{n} |w(B_i)|,$$

where the supremum is taken over the set of all finite partitions $B = (B_1, B_2, ..., B_n)$ of X. Second, the topology of weak convergence in which we say $w_t \to w^*$ weakly if for all bounded, continuous, real valued, functions ϕ on X we have that

$$\int_X \phi(x) \, w_t(dx) \to \int_X \phi(x) \, w^*(dx).$$

Note that convergence in the total variation norm implies convergence in the topology of weak convergence but not vice versa. In fact convergence

in the total variation norm implies convergence of the integrals for any measurable function ϕ. Whittle [29] contains a discussion of the relationships between the different notions of convergence for probability measures.

Assumption 4. The mapping g, from actions to probability measures over outcomes, is continuous in the total variation norm.

Assumption 4 says that if two actions are close the probability distributions over their results are similar. To see how strong this is consider a outcome mapping which produces a continuously varying outcome; this may fail Assumption 4 since even if two outcomes are close each represents an atom of probability on different outcomes. However Assumption 4 holds immediately if actions are mixed strategies, the action space has the total variation norm, and outcomes are simply the resulting pure strategy: That is, if A is the same as $\Omega(S)$ and g is the identity map. Note that in this example we require A to be endowed with the total variation norm; if it has only the topology of weak convergence continuity in the sense of Assumption 4 does not hold. Two other examples in which Assumption 4 holds are, when the set of actions is finite, or the outcome depends continuously on the actions with the addition of a independent noise term which has a continuous probability density function over S.

In order to avoid confusion we shall write the probability of the set $B \in S^*$, given the action a, generated by probability measure $g(a)$, as $g(a, B)$ rather than $(g(a))(B)$.

The history determines what is in the minds of the players when they plan their actions. Given a history h_t (in the appropriate space H which depends on whether we are applying Assumption 2a or 2b we write

$$Q(h_t, B) = \int_A g(a, B)\, \theta_{h_t}(da), \qquad h_t \in H, \ B \in S^*$$

as the probability, given a history h_t, that agents actions lead to an outcome in the (measurable) subset B of S. Given the history h_t the planned action is $f(h_t)$. The possibility of mistakes generates a probability measure $\theta(f(h_t)) = \theta_{h_t}$ over actual actions. We integrate over actual actions to get the distribution of outcomes that the history generates.

LEMMA 1. $Q: H \times S^* \to [0, 1]$ *is a stochastic kernel; that is* (*see Feller* [11]), *if*

(i) *for fixed* $h \in H$ *we have that* $Q(h, \cdot)$ *is a probability distribution over S.*

(ii) *for fixed* $B \in S^*$ *we have that* $Q(\cdot, B)$ *is a Baire function on* H.

Further, Q, when regarded as a mapping from histories to probability measures over outcomes, is continuous in the total variation norm.

Proof. See Appendix.

Lemma 1 ensures that for each possible history our mapping Q generates a well-defined probability measure on the set of possible outcomes, and does so in a continuous way.

Given an outcome s_t at time t we can define the history, in the finite memory case, at time $t+1$ by setting

$$h_{t+1} = (s_t, s_{t-1}, ..., s_{t-k+1}) = z(h_t, s_t),$$

where $z: H \times S \to H$ is clearly continuous in both h_t and s_t. In the case of infinite memory we have

$$h_{t+1} = (s_t, s_{t-1}, s_{t-2}, ...) = z(h_t, s_t)$$

which again is continuous in the product topology.

Taking h_t to be the state of the system at time t the system evolves in the following way. Given h_t the mapping Q generates a probability distribution over outcomes at time t. Given the actual realization of the outcome the updating rule z then gives us the new state at time $t+1$.

Let $D \subseteq H \times S$ and for $h \in H$ define $D_h = \{s \in S : (h, s) \in D\}$. Let H^* be the Borel subsets of H and define a function $P: H \times H^* \to [0, 1]$ by

$$P(h, G) = Q(h, (z^{-1}G)_h), \qquad h \in H, \quad G \subseteq H^*$$

$P(h, G)$ tells us the probability of going from a history h at time t to a history in the set G of histories at time $t+1$. This is just the probability of an outcome at time t in the set $(z^{-1}G)_h$, the set of outcomes which causes the history to move from h to an element of G.

LEMMA 2. *P is a transition probability on (H, H^*).*

By definition and Lemma 1 our system satisfies all the conditions required by Futia [16, Theorem 5.2] and the result follows.

Lemma 2 ensures that the probability operator P is well defined.

At any time t the state of the system is h_t. Given the stochastic evolution of the system we will be unsure of the state of the system at future times. Let w_t be a probability easure on H at time t representing our uncertainty as to the state of the system. Then we write

$$Tw_t(G) = \int_H P(h, G) \, w_t(dh).$$

The probability that the system is in a state in the set G at time $t + 1$ is given by $Tw_t(G)$. It is the probability that the system is in state h at time t and goes to an element of G from h, integrated over the possible values of h weighted by the probability of being in state h at time t. Since the equation holds for each $G \subseteq H$ (assuming $G \in H^*$) we have that Tw_t is a probability measure on H and, by its definition, it follows that

$$w_{t+1} = Tw_t.$$

Given a probability measure on the state at time t the operator T generates a probability measure on the state at time $t + 1$. Clearly T defines a Markov operator on the set of probabilities. Given an initial measure w_0 on the state of the system at time $t = 0$ (which may be an atom on a particular state h) we can write $w_t = T^t w_0$ where T^t is defined iteratively by $T^t w_0 = T(T^{t-1} w_0)$.

THEOREM 3a. *Under Assumptions* 1, 2a, 3, *and* 4 *there exists a unique invariant measure* w^* *over* H *such that* $Tw^* = w^*$ *and for all initial conditions* w_0 *we have* $w_t \to w^*$, *where convergence is in the total variation norm and occurs at a geometric rate. Further, the empirical distributions of the realizations of actions* a_t *and outcomes* s_t *converge, with probability one, to distributions which are independent of initial conditions, and depend only on* w^*.

Proof. See Appendix.

THEOREM 3b. *Theorem* 3a *holds if we replace Assumption* 2a *with Assumption* 2b.

Proof. See Appendix.

The main result in Theorem 3 is that if we draw histograms of the realizations of actions and outcomes the results, as time passes, will converge to unique distributions which are independent of the initial conditions.

We end this section by illustrating why the result holds. To illustrate the power of the result we work with a "difficult" case in which the system has an inherent tendency to unbounded behavior. Consider a simple model with one agent with one period memory. Suppose the outcome at date t is simply the action he takes so $s_t = a_t$ and with one period memory we can write

$$h_t = a_{t-1}.$$

Now suppose that agent tries to choose a_t so as to equal his expected value of s_t plus one and that he uses the simple forecasting rule $G(s_t | h_t) = h_t$;

that is, his "guess" of what will happen next period is simply what happened this period. The agent chooses $a_t = h_t + 1$; substituting for h_t we have the dynamic system

$$a_t = a_{t-1} + 1.$$

Clearly this system has no equilibrium distribution of actions; indeed it does not even have a rational expectations equilibrium. If $a_0 = 0$ we have $a_t = t$; the action keeps increasing indefinitely.

In order to meet the conditions for our theorem we need to add the possibility of mistakes. Suppose that, with probability $p > 0$, the agent makes the mistake of playing 0, his "default" action. It is now easy to show that the model satisfies Assumptions 1–4 (we use the trivial metric and discrete topology on A and S) and an equilibrium distribution of outcomes exists. It is easy to check that the long run equilibrium distribution of actions (and histories) is

$$w^*(a) = p(1 - p)^a \qquad \text{for} \quad a \geqslant 0.$$

In the long run the probability of $a = 0$ is p (the probability of a mistake) while the probability of $a = x$ is $(1 - p)$ times the probability of $a = x - 1$ ($a = x$ occurs if $a = x - 1$ occurs and the player does not make a mistake).

Our main point is that in this example the empirical distribution of actions converge to w^* for any initial condition. However, it is also worth noting that with the possibility of mistakes the model does have a rational expectations equilibrium, a^* which is $a^* = (1 - p)/p$. At a^* the probability $(1 - p)$ of going to $a^* + 1$ and the probability p of going to zero just balance to give an expected value of a^*. It is easy to check that $a^* = (1 - p)/p$ is also the expected value of a given the distribution w^*; it is therefore the limit (with probability 1), as time passes, of the empirical average of actions taken. In the long run average behavior coincides with the rational expectations equilibrium.

The example just given is set in a simple one person, one period memory, framework but it does bring out how Theorem 3 works. The key to the result is the possibility of mistakes. These ensure that the actions do not to explode to infinity. With unbounded memory a few mistakes would not matter very much but with bounded memory they must eventually dominate the explosive part of the model.

3. Expected Behavior

In general, the probability measure w^* on histories generates equilibrium probability measures on the set of actions and the set of outcomes. These

distributions represent a complete description of the long run average behavior of the system. However, in practice we often wish to condense the information in the probability distribution to a "representative" number, an expected value; this loses information but is a more parsimonious representation of the outcome.

Formally, let $\alpha: A \to R$ be a measurable function then we take the expected value of $\alpha(a)$ given a history h to be

$$E(\alpha(a)\,|\,h) = \int_A \alpha(a)\,\theta_h(da).$$

The history h generates a distribution θ_h of possible actions; we integrate over these actions to get the expected value. Then the expected value of $\alpha(a)$ given the distribution w^* on h is

$$E(\alpha(a)\,|\,w^*) = \int_H E(\alpha(a)\,|\,h)\,w^*(dh).$$

We can define $E(\beta(s))$, for arbitrary measurable functions $\beta: S \to R$, in a similar manner. For each h we have

$$E(\beta(s)\,|\,h) = \int_S \beta(s)\,Q(h,\,ds),$$

where $Q(h, s)$ is the probability of an outcome s occurring given the state h. Then to find the expected value of $\beta(s)$ we integrate over possible states h weighted by their probability:

$$E(\beta(s)\,|\,w^*) = \int_H E(\beta(s)\,|\,h)\,w^*(dh).$$

We now consider the relationship between the actual realizations of the actions and outcomes and the expected value given the equilibrium distribution.

THEOREM 4. *Under the conditions of Theorem 3a, or 3b, for any measurable real valued functions α and β, the average of the actual realizations $\sum \alpha(a_t)/T$ and $\sum \beta(s_t)/T$, $(t = 0, 1, ..., T-1)$, converge, as T tends to infinity, with probability one, to $E(\alpha(a)\,|\,w^*)$ and $E(\beta(s)\,|\,w^*)$, respectively.*

Proof. See Appendix.

The distribution of actions and outcomes is a complete picture of the long run average behavior. If we took the expected values of all functions over A and S these expected values would capture all the information in

the distribution (e.g., use indicator functions on each possible subset). By taking one, or a small number, of the expected values we may be able to characterize the average behavior of the system sufficiently for our purposes.

4. The Equilibrium Distribution

Since the equilibrium distribution w^* completely determines the long run behavior in the model it would be nice to give a simple characterization of this distribution. From the definition of T we have that the measure w^* satisfies

$$w^*(G) = \int_H P(h, G)\, w^*(dh)$$

for each measurable set G in H^*. In the case where the set of possible histories H is finite we can regard w^* as a vector of probabilities on the finite elements of H. We then have

$$w^*(h') = \sum_{h \in H} P(h, h')\, w^*(h)$$

for each h'. Letting M be the matrix of transition probabilities (with elements $P(h, h')$) we have $w^* = Mw^*$ and so $(M - I)\, w^* = 0$ and the equilibrium probabilities on the histories are given by the (unique) eigenvector of M associated with the unit eigenvalue.

It follows in the case that the set of possible histories is finite that knowledge of the transition probabilities allows the calculation of the equilibrium distribution directly; this is done explicitly in Section 6 for a simple example. In the more general case of an infinite set of possible histories the calculation of w^* relies on some insight into the structure of the particular model being examined. In practice, in such cases, it is often easiest to use computer simulation to construct empirical distributions of actions and outcomes and use Theorem 3 to argue that these approximate the equilibrium distribution.

It is interesting to compare how adding mistakes changes the nature of the equilibrium distribution in relation to the behavior of the model without mistakes. With small probabilities of mistakes we can calculate the unique invariant distribution. As the probability of mistakes approaches zero this equilibrium distribution varies and we want to know under what circumstances the equilibrium distribution will converge to an equilibrium distribution for the model without mistakes.

To investigate this consider the family of models generated by taking

different values of p, the probability that agents make a mistake. Let T_p be the operator generated by using the mistake probability p. T_0 represents Markov operator for the model in which agents do not make mistakes. For each $p > 0$ Theorem 3 guarantees that we can find a unique equilibrium distribution of histories w_p^* and that this determines the long run behavior of the system for any initial conditions.

We would like to be able to say that for p sufficiently small the equilibrium distribution w_p^* is close to an equilibrium distribution w_0^* in the model without mistakes. However, our simple example in Section 2 shows that this cannot be the case in general; in that example the model without mistakes did not even have an equilibrium distribution and the limit of the sequence of equilibria as p tends to zero is not defined. As the next theorem shows this problem does not arise if the set of outcomes, S, is compact.

THEOREM 5. *If S is compact then (under either Assumptions 2a or 2b, and 1, 3, and 4) the equilibrium correspondence, defined by $E(p) = \{w_p^*: T_p(w_p^*) = w_p^*\}$ is upper hemi continuous, in the topology of weak convergence, at $p = 0$.*

Proof. See Appendix.

Note that previously we have always used the total variation norm as our notion of distance between probability measures. Here we use a weaker notion; the sequence w_p^* has cluster points which converge weakly in distribution to the fixed points of T_0. Theorem 5 implies that if S is compact and $\phi: H \to R$ is continuous and then for any $\varepsilon > 0$ there exists $p(\varepsilon) > 0$ such that $0 < p < p(\varepsilon)$ implies that

$$\left| \int \phi(h) \, w_p^*(dh) - \int \phi(h) \, w_0^*(dh) \right| \leqslant \varepsilon$$

for some w_0^* satisfying $T_0 w_0^* = w_0^*$. To see the relationship between this result and Theorem 3 note that Theorem 3 held for any measurable function. It is easy to see that we can prove a similar result, using Theorem 5, giving convergence (as p tends to zero) for the expected values $E(\alpha(a) | w_p^*)$ and $E(\beta(a) | w_p^*)$ to $E(\alpha(a) | w_0^*)$ and $E(\beta(a) | w_0^*)$ respectively, only when α and β are continuous functions.

For $p = 0$, Theorem 3 does not hold; we can neither guarantee a unique equilibrium or global convergence. However, Theorem 5 tells us, for compact S, there is at least one invariant probability distribution in the model without mistakes, and that for low probabilities of making mistakes the equilibrium distribution is close (in distribution) to such an invariant probability. It follows that in models without mistakes that have a unique invariant distribution we can argue that adding mistakes gives rise to a

globally convergent system with behavior close to that of the unperturbed system. In particular if the unique equilibrium distribution in the unperturbed model is an atom of weight on a particular outcome we can argue that with mistakes this outcome becomes globally stable in the sense that the actual outcomes are "almost always" close to it.

Theorem 5 makes it clear that adding mistakes to the model generates a refinement of the equilibrium set. In Section 6 we give an example with multiple equilibria in the case of no mistakes and show how the presence of mistakes can be used to pick out a particular equilibrium. While mistakes are a refinement they do not necessarily pick out a unique equilibrium; in some cases the distribution of the mistakes, which actions are chosen if a mistake is made, may affect which equilibrium is selected.

While it is true in some cases that the mistake distribution matters often it does not, so long as it has full support. The intuition for this surprising result is that we can regard each equilibrium as an attractor for some neighborhood of initial conditions. Without mistakes which equilibrium is chosen depends on the initial conditions. The mistakes, however, mean that it is possible to move from one equilibrium, cross the boundary of the attracting region, and enter another equilibrium. The stability of an equilibrium with mistakes depends on how difficult it is to do this. Suppose that no move from equilibrium A to equilibrium B requires n mistakes each with a probability b in the mistake distribution; this has total probability $(pb)^n$. The corresponding probability of moving from B to the domain of attraction of A is $(pa)^m$. The relative stability of B to A depends on the ratio of these two numbers; that is, on $p^{n-m}b^n a^{-m}$. However, as p gets small this goes to infinity for $n > m$ and to zero for $n < m$. It follows that which equilibrium is chosen depends only on the relative number of mistakes needed to move from one basin of attraction to another, not on the probabilities of each type of mistake. Only in the case where the number of mistakes required to move in each direction is the same does the mistake probability matter when p becomes small. An example of this phenomenon is given in Section 6. A full theoretical treatment of this property can be found in Freidlin and Wenzel [13] while Kandori, Mailath, and Rob [19] and Young [27] show how useful it is when applied to learning and equilibrium selection in games.

5. AVERAGING OVER "VILLAGES"

So far we have discussed the average of actions and outcomes over time; it would be nice if we could say something about behavior at a point in time, for dates sufficiently far in the future. If we consider the model then, in the limit as time, t, increases, the probability distribution over the out-

come converges to the unique invariant distribution, independently of initial conditions; however, the actual outcome at time t is random.

If we were to start several identical versions of the model ("villages"), each with a different initial condition, and with different realizations of any random variables, then our probability distribution over each model's state would approach, as time passed, this same limit distribution. It follows that for large times the empirical distribution of states, taken over a large number of "villages" will approximate the limit distribution. In addition the expected value of any function, taken over the villages, will be the limit expected value. By the law of large numbers, the actual realization of the average outcome will be close to this expected value, providing we average over enough villages.

Let us suppose that our villages are indexed by j and suppose that the initial conditions in each village are independent draws from a probability distribution over H. Let $a_t(j)$ be the vector of actions taken at time t in village j. Let

$$\alpha_t = \plim_{J \to \infty} \sum_{j=1}^{J} \alpha(a_t(j))/J$$

be the average behavior of $\alpha(a_t)$ across villages. Since the $\alpha(a_t(j))$ are a sequence of independent, identically distributed, random variables the strong law of large numbers implies that this average converges with probability one (convergence in plim).

THEOREM 6. *Under Assumptions 1 to 4 we have* $\alpha_t \to E(\alpha(a) \,|\, w^*)$.

Proof. See Appendix.

If we have a large number of villages the average over these, at time t, will almost certainly be close to α_t. By Theorem 6 α_t is close to the fixed value $E(\alpha(a) \,|\, w^*)$ for t large. It follows that for a large, but finite, number of villages the average behavior will fluctuate but will almost always be close to $E(\alpha(a) \,|\, w^*)$.

This result is a proof, for our model, of the ergodic hypothesis, an hypothesis which is frequently used as an axiom in statistical mechanics. The average behavior of our system, averaging over time, is equal to the average at a point in time of the behavior in many independent systems.

To give an example consider the simple model used in Section 2 where the agent planned action is his previous action plus one, and a mistake is to play zero. If we consider many such agents, each independently following this rule, we see that averaging over the population will give a sequence of observations which converges, as time passes, to a steady state which is the expected value of the action given the equilibrium distribution.

However, we have already shown that this expected value is the rational expectations equilibrium of the model, so we have global convergence to rational expectations equilibrium.

6. APPLICATIONS

Consider a group made up of two pools of agents who play the same 2 player, normal form, finite strategy, game repeatedly, the pairing of agents in each period (one from each pool) being random. This situation has been investigated in Canning [6] under the assumptions that the set of agents is infinite and the agents' memories are bounded. Here we investigate the situation where the number of agents is finite (n) and agents sometimes make a mistake in their strategy choice (a trembling hand). Let us suppose that agents use the empirical average of their past k (remembered) observations to form their beliefs about what will happen in the next period and choose a strategy which is a best reply to this subjective belief. If agents are indifferent over a set of strategy choices we allow them to choose their action function, f_i, at this point arbitrarily from the best reply set. If agents make a mistake we assume they play a fixed mixed strategy.

Formally, we can think of a history h as consisting on a $n \times k$ matrix, giving the last k pure strategy observations of each of the n agents. H is finite since there are only finitely many such matrices (elements of the matrix are pure strategies). The set of actions is a mixed strategy for each agent; we endow each action space A_i with the total variation norm. An agent's subjective belief generates a planned action, and the possibility of a mistake (fixed probability p each period) gives a probability distribution over possible actions. Note that while each agent plans his action on the basis of the current history he actually only uses the history of his own observations to work out what to do. While we have assumed that agents all condition on the same history we can allow for asymmetric information by changing the functions f_i. The outcome of an action is an n vector assigning to each agent a pure strategy observation given by the actual strategy played by the opponent they were matched with. This is just a permutation of the actions of the agents in the other pool; under random matching we have that each possible permutation is equally likely.

PROPOSITION 7. *This model satisfies Assumptions 1 to 4.*

Proof. Since the action and outcome spaces are finite Assumption 1 obviously holds. We have assumed that agents base their actions on a finite history; the continuity necessary to secure 2 follows from the fact that the set of histories, H, is finite, so all functions on it are continuous (we take

the discrete topology on H and the usual Euclidean topology on the mixed strategy space). Assumption 3 is satisfied since we do have a positive mistake probability; the support of the distribution of mistakes is finite since the set of pure strategies is finite and the mistake distribution is over these pure strategies. Actions are probability measures over pure strategy choices and this, together with the mistakes, induces a probability measure over outcomes; the strategies actually played. Since the strategy set is finite the total variation norm on mixed strategies is the same as the Euclidean norm and Assumption 4 is satisfied.

In fact in this simple model the set of possible histories is finite; so, if the mistakes have full support on the pure strategies of the agents, the system is a finite irreducible Markov chain and we can apply elementary matrix theory to derive the ergodic result (e.g., see Cox and Miller [7]). All our continuity assumptions (Assumptions 2 and 4) hold automatically because of the finite nature of the outcome and history spaces.

To work out a simple example consider a simple coordination game. The payoff matrix is given by

		Player 2	
		A	B
Player 1	A	3, 3	0, 0
	B	0, 0	1, 1

Suppose that we have one agent in each pool and they each have one period memory, using the last action of the other player as the prediction of next period's behavior. Without mistakes three things can happen, we either get "stuck" at the strategy pair (A, A), or at (B, B), or the system oscillates giving (A, B) and (B, A) alternatively; what happens depends on in which state the system starts. As we increase the length of the agents' memories more complex things can happen (in particular it is interesting to study what happens as memories get longer and the number of agents grows). Now let us suppose we have a probability $p = 0.1$ than agents make a mistake, in which case they put equal weight (probability 1/2) on each pure strategy choice. There are four possible states for the system (A, A), (A, B), (B, A), and (B, B). Labeling these as histories 1 to 4 we can take the distribution over states at time t as w_t, where w_t is a vector with 4 elements denoting the weights on states 1 to 4. Taking the transition probabilities $p(i, j)$ to be the probability of going from state i to state j we can write $w_{t+1} = Mw_t$, were the matrix M of transition probabilities (where $M_{ij} = P(i, j)$) is given by

$$
\begin{matrix}
0.9025 & 0.0475 & 0.0475 & 0.0025 \\
0.0475 & 0.0025 & 0.9025 & 0.0475 \\
0.0475 & 0.9025 & 0.0025 & 0.0475 \\
0.0025 & 0.0475 & 0.0475 & 0.9025
\end{matrix}
$$

as compared with the matrix

$$
\begin{matrix}
1.0 & 0 & 0 & 0 \\
0 & 0 & 1.0 & 0 \\
0 & 1.0 & 0 & 0 \\
0 & 0 & 0 & 1.0
\end{matrix}
$$

in the model without mistakes. In the case where mistakes occur it is easy to see that the unique equilibrium distribution of histories puts equal weight (1/4) on each possibility. This implies that, on the average, agents play both their strategies equally often.[1]

In fact for any probability p of making mistakes the equilibrium distribution in the model is to put weight 1/4 on each possible outcome. Note that this distribution is also, by Theorem 5, an invariant probability for the case with no mistakes. If we start by putting weight 1/4 on each possible state our distribution over the states in the next period is the same. It is easy to check that this is indeed the case in our example. However, as we have already seen, in the long run we cannot have each outcome played 1/4 of the time; without mistakes we either settle into one of the coordinated equilibria or oscillate, the agents always being out of synchronization. The reason for this is that in the case of no mistakes the weights 1/4 on each outcome do give an equilibrium distribution (as indicated by Theorem 5), but Theorem 4 does not hold for $p = 0$. Without mistakes there is no guarantee that a particular equilibrium distribution determines the long run average behavior of the system.

To elaborate on this example now consider a set of n agents who play the game against the entire population each period. Each period every agent chooses either action A or B so the outcome space is an n vector of actions; that is $S = \{A, B\}^n$. The payoff to agent i from playing strategy A is given by 3 whenever he meets another "A" player and zero when he meets a "B" player. This gives an average payoff of 3α, where α is the

---

[1] It is interesting to note that the matrix M is symmetric; $p(i, j) = p(j, i)$. This is a special case rather than usual in models of the type we are describing, but is axiomatic in statistical mechanics when applied to physics. It follows, in particular, that in our example we can derive the law of increasing entropy; a large collection of pairs of agents, each acting as in our model, will tend to become more disordered as time passes (see Waldram [28]). The limit distribution puts equal weight on each possible state, which is the most disorder that is possible.

fraction of the population playing strategy A. The payoff to strategy B is $(1 - \alpha)$. Note that we assume that the agent plays himself when calculating these payoffs. With complete one period memory agents can calculate last period's value of α. An alternative story would be that agents only meet pairwise each period but a central agency announces the value of α. If agents assume that α is stationary and maximize their expected payoff given last period's value all agents will try to take the same action. They will play A if $3\alpha > (1 - \alpha)$ and B if $3\alpha < (1 - \alpha)$. The critical value of α is $1/4$, above this agent play A below this they play B. We assume that N is chosen so that $\alpha = 1/4$ is not possible to avoid the problem of indifference between actions.

Clearly both $\alpha = 1$ and $\alpha = 0$ are steady states of the system without mistakes. The only equilibrium distributions of this system are weights which sum to one on the two states $A^* = (A, A, A, ..., A)$ and $B^* = (B, B, B, ..., B)$. Suppose we add the possibility of mistakes. If a mistake always leads to playing A then A^* will be selected as the unique equilibrium. If a mistake always leads to a B being played then B^* will be the unique equilibrium. The more interesting case is when a mistake gives a positive probability to both A and B being played; let a be the probability on A and b the probability of B $(a + b = 1)$. In this case all states in S are possible even in the long run and the equilibrium distribution has full support on S ($H = S$ since we have one period memory). However, as the probability of a mistake goes to zero the weight on states other than A^* and B^* must go to zero. Let

$$S_A = \{s \in S \mid \alpha(s) > \tfrac{1}{4}\}, \qquad S_B = \{s \in S \mid \alpha(s) < \tfrac{1}{4}\}.$$

In the equilibrium distribution the probability of entering any set of histories must equal the probability of leaving it so that

$$P(S_A \rightarrow S_B)\, w^*(S_A) = P(S_B \rightarrow S_A)\, w^*(S_B)$$

since the probability of leaving the set S_A or S_B is independent of which actual state in those sets the system is in. In S_A all agents will try to play A; in order to leave S_A we require at least $z = 3N/4$ agents to make mistakes and play B. Similarly to leave S_B at least $N/4$ agents must make mistakes and play A. It follows that

$$P(S_A \rightarrow S_B) = \sum_{i=z}^{N} \frac{N!}{(N-i)!\, i!} (pb)^i (1 - pb)^{N-i}$$

$$P(S_B \rightarrow S_A) = \sum_{i=N-z+1}^{N} \frac{N!}{(N-i)!\, i!} (pa)^i (1 - pa)^{N-i},$$

where z is the smallest integer greater than $3N/4$. Hence

$$\frac{w^*(S_B)}{w^*(S_A)} = \left(\frac{b}{a}\right)^{N-z+1} \frac{\sum_{i=z}^{N}(N!/(N-i)!\,i!)(pb)^{i-(N-z+1)}(1-pb)^{N-i}}{\sum_{i=N-z+1}^{N}(N!/(N-i)!\,i!)(pa)^{i-(N-z+1)}(1-pa)^{N-i}}.$$

Taking limits as p tends to zero it is clear that every term on the top line contains a power of p while the bottom line has a first term which is strictly bounded away from zero as p goes to zero; the key is that we have $z > (N+1)/2$. Hence the ratio $w^*(S_B)/w^*(S_A)$ goes to zero as p tends to zero. It follows that as the probability of a mistake is reduced the weight on the set S_A under w^* goes to one. Since the only equilibrium distribution in this set for $p = 0$ is weight one on the state A^*. It follows from Theorem 5 that

$$\lim_{p \to 0} w_p^*(A^*) = 1.$$

With a small probability of making a mistake the state will almost always be in the state $A^* = (A, A, A, ..., A)$. Only the states A^* and B^* can be persistent; any other state usually leads immediately to one of these. The system can get stuck in the state B^*, but it takes only $N/4$ mistakes to leave its "basin of attraction," S_B. It takes $3N/4$ mistakes to leave S_A the basin of attraction for A^*. It follows that A^* is more stable than B^* under "small" random perturbations.[2]

For any symmetric game where the payoffs of one equilibrium risk dominate (in the sense of Harsanyi and Selten [18]) those of the other it is easy to show, using the method above, that the one period memory system will select the risk dominant equilibrium, as the probability of a mistake goes to zero. When the two equilibria rank equally (e.g., if we replace the payoff $(3, 3)$ with the payoff $(1, 1)$ in the example above) the long run behavior of the system (for p small) approximates a correlated equilibrium with half the time spent in one Nash equilibrium and half in the other. Di Gioacchino [8] derives similar results in simulations of two players with memory of each others past actions that extends more than one period. Kandori, Mailath, and Rob [19] also have the same result in a class of models with one period memory in which a proportion of agents adapt each period.

Now consider a more indirect application of our results. Suppose in each period there is a finite set E of events. Suppose agents believe that the

---

[2] The probability of a perturbation is small as p becomes small. However, the perturbations can, though with low probability, be big. It is these low probability, large perturbations, which allow us to perform a global (independent of initial conditions) stability analysis as opposed to the local stability analysis usually associated with small perturbations.

events are independent, identically distributed, random variables, but are unsure of their distribution. Note that the belief that the variables are stationary need not be correct. Let us suppose they begin with a prior distribution over the space of probability distributions on events which has full support and update this belief with their experience. The Dirichlet functions form a conjugate class for this problem; if the prior belief is in the class so is an agent's posterior given his observation. The Dirichlet functions can be parameterized by a finite vector of real numbers; it follows, if agents' priors lie in this class, that we can represent the agent's belief by a vector drawn from a finite dimensional Euclidean space.

Let us suppose the actual event which occurs is a random variable which depends on the agents's actions and has a distribution which is continuous (in the total variation norm). We take events to be continuous in actions which are continuous in beliefs. Given their beliefs agents decide on actions which give rise to events which cause them to update their beliefs. In addition we suppose that with probability p each period the agent dies and is replaced by a newborn agent with a fixed prior drawn from a (finite support) distribution on the set of Dirichlet functions. Note that this formulation is exactly that used by Fudenburg and Levine [15] so our result will apply to their model.

In order to apply our results to this model let $\Omega(E)$ be the set of probability measures on E and let D be the set of Dirichlet functions (each a probability distribution on $\Omega(E)$). Suppose $S = D^n$, where n is the number of agents and let agents have one period memory so $H = S$. Let $A = S$ and suppose each agent's action is simply to state his belief. Given this choice of action we perturb it, with probability p replacing it with the random draw from the newborn. We then map the vector of actual actions (given the outcome of the perturbation) into a probability easure over outcomes by finding the outcome for each possible event e (the updating of beliefs that results) and weighting them by the probability of e given the action taken. An outcome in S is simply a new belief for each agent, which depends on their old belief and the event which occurs. Note that instead of allowing the agents to take an action directly we use as their "action" a statement of their belief, derive the action they would have taken from this, and take their new belief as the outcome.

PROPOSITION 8. *Our overlapping generations model satisfies Assumptions 1–4.*

Proof. A_i and S are finite dimensional Euclidean spaces so Assumption 1 holds. Assumption 2 holds trivially since planned actions are equal to last period's outcome. Assumption 3 holds because of the random deaths and replacement. Assumption 4 holds since the mapping from actions to

events is assumed continuous in the total variation norm and it is easy to show that the updating rule is continuous in beliefs for fixed events.

In an overlapping generations model we need not assume bounded memory or mistakes. The possibility of death effectively bounds memory and the newborn, having randomly drawn prior beliefs, can be considered as making "mistakes." It follows that in a macroeconomic model with over-lapping generations we can derive convergence to a unique equilibrium for the time average of behavior, and even, if the economy is made up of enough independent "villages" convergence of the macroeconomic variables themselves. We may have persistent, stable, average behavior, despite the fact that agents' decision rules need not be "rational." This gives rise to the possibility that, as in physics, average behavior, averaging over a large number of agents (or particles), is better behaved, and more predictable, than individual behavior.

7. Conclusion

We have examined a wide class of learning models in which agents inter-act and shown that provided we assume bounded memory and mistakes the average behavior in a model will converge to a long run limit that is unique and independent of the initial conditions. The power of the result lies in the fact that it applies to any adaptive behavior rule in any model of social interaction which satisfies some weak conditions (notably con-tinuity of actions with respect to histories and continuity of the probability measures over outcomes with respect to actions). While we prove con-vergence and uniqueness results we leave open the characterization of the average behavior; the nature of the equilibrium depends on the behavioral rules that agents follow.

Appendix

LEMMA 1. $Q: H \times S^* \to [0, 1]$ is a stochastic kernel; that is (see Feller [11]) if

(i) for fixed $h \in H$ we have that $Q(h, \cdot)$ is a probability distribution over S.

(ii) for fixed $B \in S^*$ we have that $Q(\cdot, B)$ is a Baire function on H.

Further, Q, when regarded as a mapping from histories to probability measures over outcomes, is continuous in the total variation norm.

Proof. For (i) we need to show that $Q(h, \cdot)$ defines a probability measure, but this follows immediately from the fact that for each a the function $g(a, \cdot)$ defines a probability measure on S and θ_h defines a probability measure on A. To check, it is easy to see that

$$Q(h, B) = \int_A g(a, B)\, \theta_h(da), \qquad h \in H, \ B \in S^*$$

defines a measure on S and that $Q(h, S) = 1$.

For (ii) it is sufficient to show that $Q(\cdot, B)$ is continuous, since the Baire functions are the closure of the set of continuous functions. Without mistakes the probability measure θ_h is just an atom of weight 1 on the action $f(h)$. Therefore, in the absence of mistakes, we can write

$$|Q(h, B) - Q(h', B)| = |g(f(h), B) - g(f(h'), B)|,$$

which, since both g and f are continuous (by Assumptions 2 and 4) implies $Q(\cdot, B)$ is continuous. With mistakes we can decompose the probability measure θ_h into 2^n possible cases, depending on which agents make mistakes and which do not. Hence we can write

$$|Q(h, B) - Q(h', B)| = \left| \sum_m g(a_m(h), B)\, p(m) - \sum_m g(a_m(h'), B)\, p(m) \right|$$

$$\leqslant \sum_m |g(a_m(h), B) - g(a_m(h'), B)|\, p(m),$$

where the subscript m takes 2^n values, $p(m)$ is the probability of the combination of mistakes, m, and the actions $a_m(h)$ and $a_m(h')$ refer to the actions taken given the combination of mistakes, m (assuming a mistake leads to a single choice of action).

In the case where the agent chooses from a probability distribution over actions (rather than select a single action) when he makes a mistake we must, for each combination of mistakes m, consider the possible actions agents take. However, since we have assumed finite supports, and the probability distributions over these supports when a mistake occurs is history independent, we can write

$$|Q(h, B) - Q(h', B)| = \left| \sum_m \sum_c g(a_{mc}(h), B)\, p(m)\, p(c) \right.$$

$$\left. - \sum_m \sum_c g(a_{mc}(h'), B)\, p(m)\, p(c) \right|$$

$$\leqslant \sum_m \sum_c |g(a_{mc}(h), B) - g(a_{mc}(h'), B)|\, p(m)\, p(c),$$

where c is an index over the set of all possible mistakes the agents can make. If an agent makes a mistake in the combination m, then his action is given by the choice in c. If he does not make a mistake c is irrelevant for him and he takes his planned action which depends on h. It is easy to see that each a_{mc} is continuous in h and the continuity of Q follows.

THEOREM 3a. *Under Assumptions* 1, 2a, 3, *and* 4 *there exists a unique invariant measure* w^* *over* H *such that* $Tw^* = w^*$ *and for all initial conditions,* w_0, *we have* $w_t \to w^*$, *where convergence is in the total variation norm and occurs at a geometric rate. Further, the empirical distributions of the realizations of actions* a_t *and outcomes* s_t *converge, with probability one, to distributions which are independent of initial conditions, and depend only on* w^*.

Proof. Our proof has four parts. First, in (i) we show that the operator T is quasi-compact. In (ii) we show that its adjoint is stable (in the sense of mapping continuous functions into continuous functions) which implies that T is equicontinuous and invariant probability measures, w^*, exist. Next, in (iii) we show that the transition probability P, which underlies T, satisfies a uniqueness condition which ensures that w^* is unique and that $T^t w_0 \to w^*$. Finally in (iv) we show that the empirical realizations of actions and outcomes converge, with probability one, to those generated by the unique invariant distribution on histories, w^*.

(i) By Theorem 4.9 of Futia [16] a Markov operator is quasi-compact if its transition probability satisfies Doeblin's condition. Doeblin's condition is that there exists a probability measure δ on H, an integer m, and an $0 < \varepsilon < 1$ such that if $G \in H^*$ and $\delta(G) \leqslant \varepsilon$ then $P^m(h, G) \leqslant 1 - \varepsilon$ for all h, where P^m is the mth iterate of the transition probability P.

Let δ be the probability measure generated on H in the case that all agents simultaneously make mistakes for k periods and set $m = k$, the length of agents' memories. Set $\varepsilon = p^{kn}/(1 + p^{kn})$ where n is the number of agents in the system. Suppose $\delta(G) \leqslant \varepsilon$. Now

$$P^k(h, G) \leqslant 1 - p^{nk} \delta(G');$$

the probability of arriving in the set G after k periods is bounded above by one minus the probability that everyone makes mistakes in all k periods and the result of these mistakes is to take the system into its complement G'. Since $\delta(G) \leqslant \varepsilon$ we have $\delta(G') > 1 - \varepsilon$ and so

$$P^k(h, G) \leqslant 1 - p^{kn}(1 - \varepsilon) = 1 - \varepsilon$$

by our choice of ε. Hence P satisfies Doeblin's condition and T is quasi-compact.

(ii)   Now we show that T is equicontinuous. By Futia [16, Theorem 3.3] given that T is quasi-compact, all we need to show is that its adjoint operator is stable. But Futia [16, Theorem 5.6] shows that such stability follows if the mapping Q, regarded as a mapping from histories to probability measures over outcomes, is continuous in the total variation norm. However this was shown in Lemma 1.

(iii)   Let h_0 be in the support of the probability distribution δ. Then for every neighborhood U of h_0 we have $\delta(U) > 0$ (otherwise we could find a set, excluding U, with probability 1, and exclude h_0 from the support of δ, which is just the closure of the smallest subset of H with probability 1). Further, we have $P^k(h, U) > 0$, since there is a probability $p^{nk} > 0$ that in the last k periods all n agents have made mistakes and the current history is a random draw using the probability distribution δ. It follows that the transition probability P satisfies the uniqueness criterion given by Futia [16, Theorem 2.11]. By (ii) the operator T is equicontinuous, hence by Futia [16, Theorem 2.12], there is a unique w^* satisfying $Tw^* = w^*$. Since P satisfies Futia [16, Theorem 2.12] it clearly satisfies the generalized uniqueness criterion given by Futia 3.5 and so by Futia [16, Theorems 3.6 and 3.7] we have that $T'w_0 \rightarrow w^*$, convergence being in the total variation norm, and at a geometric rate (there exists K, $\varepsilon > 0$ such that $|w_t - w^*| < K/(1 + \varepsilon)^t$).

(iv)   To show the convergence of average behavior in the model let $B \subseteq A$ be a set of actions and let $x_B(a)$ be an indicator function, taking the value one if $a \in B$ and 0 otherwise. Let $B_T = \sum x_B(a_t)/T$ $(t = 0, 1, ..., T-1)$ be the proportion of times the actual action taken is in the set B in the first T periods.

Since a_t is a random variable depending on the state h_t we can think of B_T as a random variable depending on h. Let the sequence of random variables $x_B(a_t)$ be written as $X(T'w_0)$, that is they have distribution X depending on $w_t = T'w_0$. Clearly, since each X lies between zero and one their expected values must be bounded. By Birkhoff's ergodic theorem (see Brieman [4]) the convergence of w_t to w^* implies that for any sequence of random variables such that $E(|X_t|) < \infty$ we have

$$\sum_{t=0}^{T-1} X_t/T \rightarrow E(X \mid w^*)$$

with probability 1.

THEOREM 3b.   *Theorem* 3a *holds if we replace Assumption* 2a *with Assumption* 2b.

Proof. We show that T is random contraction.

(i) Note first that since S is compact so is H (by Tychonoff's theorem, see Dunford and Schwartz [9, Theorem I.8.5]). Hence

$$\sup_{h, h'} d(h, h') < \infty$$

(ii) Let λ be the probability measure defined on (S, S^*) when all n agents make a mistake simultaneously. Clearly for any history h and measurable set of outcomes B, $Q(h, B) \geq p^n \lambda(B)$.

(iii) Let

$$r = \sup_{h \neq h'} \int_S \frac{d(z(h, s), z(h', s))}{d(h, h')} \lambda(ds),$$

where d is the metric induced by our norm on H. It is easy to see that under the updating rule z, for the same outcome s, we have that $d(z(h, s), z(h', s)) = d(h, h')/2$ and, since for $h \neq h'$ we have $d(h, h') > 0$, it follows that $r = 1/2$.

In addition note that since H is compact the operator T must be tight. It follows that T satisfies all the conditions of Futia [16, Theorems 6.11 and 6.12], and $w_t = T' w_0$ converges to the unique invariant probability measure w^* of T at a geometric rate. The rest of the proof follows from the application of Birkoff's ergodic theorem as in Theorem 3a.

THEOREM 4. *Under the conditions of Theorem 3, for any measurable real valued functions α and β, the average of the actual realizations $\sum \alpha(a_t)/T$ and $\sum \beta(s_t)/T$ ($t = 0, 1, ..., T-1$) converge, as T tends to infinity, with probability 1, to $E(\alpha(a) | w^*)$ and $E(\beta(s) | w^*)$, respectively.*

Proof. The result follows immediately from the application of Birkhoff's ergodic theorem in a similar way as in the proof of Theorem 3.

THEOREM 5. *If S is compact then (under Assumption 2a or 2b, and 1, 3, and 4) the equilibrium correspondence, $E: p \to w_p^*$, is upper hemi continuous, in the topology of weak convergence, at $p = 0$.*

Proof. We can write T_p as $T_p = (1 - p)^n T_0 + R_p$. With probability $(1 - p)^n$ no one makes a mistake and the operator acts identically to the model without mistakes. The operator R_p represents the sum of the other $2^n - 1$ operators which represent what happens under each possible combination of mistakes. Given a pattern of mistakes we can apply the appropriate operator to find the distribution over the new history. Each of

these operators has norm one in the uniform operator topology (see Dunford and Schwartz [9]) so, adding the weights on those that make up R_p we have that R_p is a linear operator with norm $1 - (1 - p)^n$. Hence

$$|T_p - T_0| \leqslant 2(1 - (1 - p)^n).$$

So $T_p \to T_0$ as $p \to 0$.

Let (p) be a sequence of positive probabilities which converge to zero. If S is compact so is $H = S^k$, for k finite, as is $H = S^\infty$ using the product topology, and it follows that the family of probability measures $\{w_p^*\}$ is tight. By Billingsley [1, Theorem 1.1], a bounded measure defined on the Borel subsets of a metric space is regular; since H is metric the family $\{w_p^*\}$ is regular. Billingsley [1, Theorem 6.1] shows that a tight set of probabilities is weakly sequentially compact where the weak topology is defined by the linear functionals $w \to \int f(h)\, w(dh)$, with f continuous, and w regular. It follows that w_p^* has a convergent subsequence, w_q^*, (in the weak topology), converging to w_0^* say.

Restricting p to the converging subsequence q we have $T_0 w_0^* = \lim T_0 w_q^* = \lim T_q w_q^* = \lim w_q^* = w_0^*$, where the limits (in the weak topology) are taken as q tends to zero. Hence, w_0^* is a fixed point of T_0.

Now suppose the equilibrium set is not upper hemi continuous (in the weak topology) at zero. Then we can find a bounded continuous function f and an $\varepsilon > 0$ such that $|\int f(h)\, w_p^*(dh) - \int f(h)\, w_0^*(dh)| > \varepsilon$, for every fixed point w_0^* of T_0, for p arbitrarily small. If this is so we can construct a sequence (p) tending to zero for which this condition holds. By weak sequential compactness the infinite sequence (w_p) has a convergent subsequence and, by the argument above, its limit, in the weak topology, is a fixed point of T_0, a contradiction. Hence the condition $|\int f(h)\, w_p^*(dh) - \int f(h)\, w_0^*(dh)| > \varepsilon$ holds only for p bounded away from zero and our result follows.

THEOREM 6. *Under Assumptions 1 to 4 we have $\alpha_t \to E(\alpha(a))$ where the expectation is taken relative to the invariant distribution w^*.*

Proof.

$$\alpha_t = \plim_{J \to \infty} \sum_{j=1}^{J} \alpha(a_t(j))/J$$

hence, by the strong law of large numbers

$$\alpha_t = E(\alpha(a)\,|\,wt) = \int_H E(\alpha(a)\,|\,h)\, w_t(dh)$$

By Theorem 3 $w_t \to w^*$ in norm, hence

$$\int_H E(\alpha(a)\,|\,h)\,wt(dh) \to \int_H E(\alpha(a)\,|\,h)\,w^*(dh) = E(\alpha(a)\,|\,w^*).$$

REFERENCES

1. P. BILLINGSLEY, "Convergence of Probability Measures," Wiley, New York, 1968.
2. A. BRANDENBURGER AND E. DEKEL, The role of common knowledge assumptions in game theory, in "The Economics of Missing Markets, Information and Games" (F. Hahn, Ed.), Oxford Univ. Press, Oxford, 1989.
3. M. M. BRAY AND N. E. SAVIN, Rational expectations, learning, and model specification, Econometrica 54 (1986), 1129–1160.
4. L. BRIEMAN, "Probability," Addison–Wesley, Reading, MA, 1964.
5. D. CANNING, Equilibrium, learning and economic dynamics, mimeo, Pembroke College, Cambridge, 1990.
6. D. CANNING, "Social Equilibrium," Economic Theory Discussion Paper, No. 150, Cambridge University, 1990.
7. D. R. COX AND H. D. MILLER, "The Theory of Stochastic Processes," Chapman & Hall, London/New York, 1987.
8. D. DI GIOACCHINO, Learning and equilibrium selection in 2 × 2 games, some simulation results, mimeo, Cambridge University, 1991.
9. N. DUNFORD AND J. T. SCHWARTZ, "Linear Operators," Part I, Interscience Publishers, New York, 1957.
10. R. EVANS, Out of equilibrium learning: Analysis based on cognitive-scientific models, mimeo, St. John's College, Cambridge, 1990.
11. W. FELLER, "An Introduction to Probability Theory," 2nd ed., Vol. II, J. Wiley, New York, 1971.
12. D. FOSTER AND P. YOUNG, Stochastic evolutionary game dynamics, Theoretical Population Biology 38 (1990), 219–232.
13. M. I. FREIDLIN AND A. D. WENZEL, "Random Perturbations of Dynamical Systems," Springer-Verlag, New York, 1984.
14. D. FUDENBURG AND D. KREPS, A theory of learning, experimentation, and equilibrium in games, mimeo, MIT, and Stanford, 1988.
15. D. FUDENBURG AND D. LEVINE, Steady state learning and self confirming equilibrium, mimeo, Department of Economics, MIT, and Department of Economics, UCLA, 1990.
16. C. A. FUTIA, Invariant distributions and the limiting behavior of Markovian economic models, Econometrica 50 (1982), 377–408.
17. J. C. HARSANYI, Games of incomplete information played by Bayesian players, Parts I, II, and III, Manage. Sci. 14 (1967/68), 159–182, 320–334, 486–502.
18. J. C. HARSANYI AND R. SELTEN, "A General Theory of Equilibrium in Games," MIT Press, Cambridge, 1988.
19. M. KANDORI, G. J. MAILATH, AND R. ROB, "Learning, Mutation, and Long Run Equilibrium in Games," University of Pennsylvania, CARESS Working paper No. 91-01, 1991.
20. J. MAYNARD SMITH, "Evolution and the Theory of Games," Cambridge Univ. Press, Cambridge, 1982.
21. M. F. NORMAN, "Markov Processes and Learning Models," Academic Press, New York/London, 1972.

22. H. SABOURIAN, "The Folk Theorem for Repeated Games with Bounded (one-period) Memory," Economic Theory Discussion Paper, No. 143, Cambridge University, 1989.
23. L. S. SAVAGE, "Foundations of Statistics," Wiley, New York, 1954.
24. R. SELTEN, Reexamination of the perfectness concept for equilibrium points in extensive games, *Int. J. Game Theory* **4** (1975), 25–55.
25. N. STOKEY AND R. LUCAS, "Recursive Methods in Economic Dynamics," Harvard Univ. Press, Cambridge, 1989.
26. H. P. YOUNG, The evolution of bargaining conventions, mimeo, School of Public Affairs, University of Maryland, 1989.
27. H. P. YOUNG, Conventional equilibria, mimeo, School of Public Affairs, University of Maryland, 1991.
28. J. R. WALDRAM, "The Theory of Thermodynamics," Cambridge Univ. Press, Cambridge, 1985.
29. P. WHITTLE, "Systems in Stochastic Equilibrium," Wiley, London, 1986.