For the Economic Theory Workshop in Honor
of Roy Radner, Cornell University, June 1992

Rational Expectations and Rational Learning

Lawrence E. Blume and David Easley†

July 1993

† Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14850. Financial support from NSF Grant SES-8921415 is gratefully acknowledged.

For the Economic Theory Workshop in Honor
of Roy Radner, Cornell University, June 1992

Rational Expectations and Rational Learning

Abstract

In this paper we provide an overview of the methods of analysis and results obtained, and, most important, an assessment of the success of rational learning dynamics in tying down limit beliefs and limit behavior. We illustrate the features common to rational or Bayesian learning in single agent, game theoretic and equilibrium frameworks. We show that rational learing is possible in each of these environments. The issue is not in whether rational learning can occur, but in what results it produces. If we assume a natural complex parameterization of the choice environment all we know is the rational learner believes that his posteriors will converge somewhere with prior probability one. Alternatively, if we, the modelers, assume the simple parameterization of the choice environment that is necessary to obtain positive results we are closing our models in the ad hoc fashion that rational learning was inroduced to avoid. We believe that a partial resolution of this conundrum is to pay more attention to how learning interacts with other dynamic forces. We show that in a simple economy, the forces of market selection can yield convergence to rational expectations equilibria even without every agent behaving as a rational learner.

Keywords: adaptive behavior, bounded rationality, learning, Nash equilibrium, rational expectations equilibrium.

Correspondent:

Professor David Easley
Department of Economics
Uris Hall
Cornell University
Ithaca, NY 14850

He that knew all that ever learning writ,
Knew only this—that he knew nothing yet.
——— Mrs. Aphra Behn
The Empress of the Moon, I.iii

1. Introduction

The issue of expectation determination arises very naturally in economies with a sequence of spot markets and incomplete futures markets. In such economies, individuals must forecast future prices in order to make decisions about current consumption and investment. Absent any structure on expectations, there is little to be said about equilibria at any date; they may even fail to exist. Roy Radner's (1972) seminal treatment of economies with a sequence of incomplete markets, fixed expectations by requiring that agents hold common price expectations and that their plans be consistent. Their expectations are thus "self-fulfilling" or "rational". The term "self-fulfilling" is particularly apt because it emphasizes that the actual sequence of prices is determined by the expectations agents use. The specification of a "self-fulfilling" model is endogenously determined. Radner shows that such equilibria exist. But it is not at all clear how these equilibria can be achieved when market participants may initially hold very diverse expectations.

Radner's (1972) model was concerned with economies in which all market participants have access to the same information. Radner also pioneered the study of competitive equilibrium when market participants are asymmetrically informed. In these economies, each trader must infer other traders' information from the market price and his own private information. To conduct the inference, each trader must have a model of the relationship between private information and prices. Again the equilibrium concept involves inference with "self-fulfilling" models. Radner (1979) demonstrates the existence of equilibrium with self-fulfilling, or rational expectations. Again, given the delicate structure of the equilibrium, the question of how such equilibria can be realized begs for an answer.

The subsequent literature has addressed the problem of equilibrium attainment, learning rational expectations, in two distinct ways. One approach directly postulates hypotheses about the learning model, and the goal of this approach is to identify those learning behaviors which lead to rational expectations. This the literature dismissively refers to as *ad hoc* learning. The second approach derives learning behavior from preferences. Specifically, if a market participant is an expected utility maximizer, then, as a consequence of this assumption, beliefs must be revised in light of new information according to Bayes rule. Because Bayesian learning is a consequence of assumptions about preferences, it is frequently referred to as *rational learning.* We will follow this practice, but the reader should keep in mind that there is nothing necessarily irrational about *ad hoc* learning. To label non-Bayesian learning as irrational is to invest the Savage axioms with normative content that most economists would reject.

Game theory presents learning issues similar to the issues of expectation formation in economies with a sequence of incomplete markets and markets with differentially informed traders. In games with incomplete information, a (Bayes-Nash) equilibrium implies that, throughout the course of play, players will be learning. But many different structures of beliefs will be consistent with many, distinctly different equilibria. Jordan (1991a, 1991b) investigates the equilibrium behavior of infinitely repeated games. He demonstrates how the effects of learning force a relationship between limit beliefs in such a game and the equilibria of complete information versions of the game. Kalai and Lehrer (1992a) and Nyarko (1991,1992) ask whether players can learn their way to a Nash equilibrium when they do not necessarily start in a Bayes-Nash equilibrium. In different, but related, models they both provide positive answers. The issues that arise in this literature are essentially the same as those which arise in the microeconomic rational expectations literature.

In this paper our goal is not to survey the work on equilibrium under uncertainty or on the existence of rational expectations equilibrium, nor even to survey all the recent work on rational learning.[1] Instead, our goal is to provide an overview of the methods of analysis and results obtained, and, most important, an assessment of the success of rational learning dynamics in tying down limit beliefs and limit behavior in game theoretic and economic equilibrium models. We illustrate the features common to rational or Bayesian learning in single agent, game theoretic and equilibrium frameworks. We show that rational learing is possible in each of these environments. The issue is not in whether rational learning can occur, but in what results it produces. If we assume a natural complex parameterization of the choice environment all we know is the rational learner believes that his posteriors will converge somewhere with prior probability one. Alternatively, if we, the modelers, assume the simple parameterization of the choice environment that is necessary to obtain positive results we are closing our models in the ad hoc fashion that rational learning was inroduced to avoid. We believe that a partial resolution of this conundrum is to pay more attention to how learning interacts with other dynamic forces. We show that in a simple economy, the forces of market selection can yield convergence to rational expectations equilibria even without every agent behaving as a rational learner.

In the next section we discuss learning in the context of a single-agent decision problem. Along the way we introduce some of the tools that have proven useful in the analysis of learning dynamics. Section 3 discusses the role of learning in the analysis of repeated games, and Section 4 discusses learning in general equilibrium models. In Section 5 we discuss the robustness of learning in equilibrium models. Our conclusions about what we have learned from the learning literature and what we need to learn are contained in

[1] Blume, Bray and Easley (1982) provide a survey of learning in economies with differential information, Blume and Easley (1993) provide a partial survey of the recent work on learning in games and Jordan (1992) provides an exposition of recent results on Bayesian learning in games and a non-Bayesian interpretation of some of these results.

Section 6.

2. Learning Dynamics

The rational learning literature takes off from the analysis of Bayesian decision problems. Here we establish the basic results for the single-agent learning problem. The problem fundamental to the statistical literature is *c*onsistency./ That is, will a decision maker ultimately learn the truth? We will introduce another problem which is important for equilibrium dynamics: The *p*rediction problem./ That is, does the prediction of the future path of the process given its history through date t converge to the correct conditional distribution as t grows. We will see that the relationship between consistency and the prediction problem is not as straightforward as it might seem. In this section we discuss the dynamics of Bayesian posterior beief revision and the problems of consistency and prediction. We then describe a canonical decision problem, and discuss the problem of incomplete learning in some examples.

2.1. The Dynamics of Posterior Revision

Bayesian posterior revision works on a set of sample histories $H = \prod_{t=1}^{\infty} H_t$, where H_t is the set of possible observations at time t, a set of parameters Θ, and for each $\theta \in \Theta$ a probability measure μ_θ on H. We assume that Θ and each H_t are Polish (coomplete, seperable metric) spaces. We let $\mathcal{S}$ denote the product σ-field of subsets of H derived from the Borel σ-fields on each H_t, and we assume that for each event $S \in \mathcal{S}$, the map $\theta \mapsto \mu_\theta(S)$ is Borel measureable.

The Bayesian "learner" begins with a prior distribution ν on Θ. Corresponding to each prior ν is the (unique) joint distribution ϕ_ν on $\Theta \times H$ such that for any set $A \times B$ with A a measurable subset of Θ and $B \in \mathcal{S}$,

$$\phi_\nu(A \times B) = \int_A \mu_\theta(B) d\nu(\theta).$$

Just as ν is the marginal distribution of ϕ_ν on Θ, let μ_ν denote the marginal distribution of ϕ_ν on H.

Posterior beliefs are just conditional distributions derived from ϕ_ν. Let $H^T = H_1 \times \cdots \times H_T$ denote the set of possible observations through time T. Given a measurable set $B^T \subset H^T$, the date $T+1$ posterior distribution assigns to each measurable subset A of Θ the conditional probability:

$$\begin{aligned} \nu_{T+1}(A \,|\, B^T) &= \phi_\nu(A \times H \,|\, B^T \times \prod_{t=T+1}^{\infty} H_t), \\ &= E\{1_{A \times H} \,|\, B^T \times \prod_{t=T+1}^{\infty} H_t\}, \end{aligned}$$

where E is the expectation operator with respect to ϕ_ν and $1_{A\times H}$ is the indicator function of $A \times H$ on $\Theta \times H$.

There are two key results on the consistency of Bayes learning, both essentially due to Doob (1949):

Theorem 2.1: Given any prior belief ν on Θ, posterior beliefs converge ν-almost surely.

This is to say, for ν-almost all θ, the posterior beliefs $\nu_{T+1}(\cdot \mid h_1, \ldots, h_T)$ converge in the weak convergence topology with μ_θ-probability 1. In other words, for most parameter values θ conceivable from the ex-ante point of view of the learner, for almost all possible realizations of the data, posterior beliefs will converge somewhere. This result is an immediate consequence of the martingale convergence theorem.

Theorem 2.1 does not imply learning. Limit posterior beliefs may not be correct. It may not be the case that for ν-almost all θ, posterior beliefs converge μ_θ-almost surely to point mass at θ, δ_θ. The second result is:

Theorem 2.2: If for ν-almost all θ, the measures μ_θ on H are mutually singular, then for ν-almost all θ, posterior beliefs will μ_θ-almost always converge to δ_θ.

In this case, Bayes learning is said to be **consistent**.

The condition that the measures μ_θ be mutually singular seems very strong, but in fact is true in many conventional statistics problems. Consider, for example, learning the mean θ of a normal random variate from independent draws. Let x_t denote the outcome of the t'th draw. According to the strong law of large numbers, the support of each μ_θ the induced measure on the infinite product space, is almost surely contained in the set of all sample paths whose sample means converge to θ:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t = \theta.$$

Thus the intersection of the supports of any two distinct parameters θ and θ' has measure 0 according to both μ_θ and $\mu_{\theta'}$.

A slightly more useful result for our purposes will account for the fact that the measures μ_θ may not all be mutually singular. Let A be a Borel subset of Θ, and let H_A denote the subset of paths $h \in H$ such that h is not in $\operatorname{supp} \mu_\theta$ for any $\theta \in A$. It is easy to see that H_A is a Borel set.

Corollary 2.1: For μ_ν-almost all $h \in H_A$, the posterior probability $\nu_{T+1}(A \mid h_1, \ldots, h_T)$ converges to 0.

There is one important issue here which we have avoided. The notion of "almost sure" for parameters is from the point of view of the individual, and not necessarily from the point of view of the modeler. Non-pathological examples are known in which the exceptional sets are very large from some other, say topological point of view. Thus, one would like to show that Bayes learning is "uniformly consistent" if posterior beliefs converge uniformly across $\theta \in \Theta$ to δ_θ with probability 1. One approach to this problem is introduced in Schwartz (1965) and extended in Barron (1989).

These learning results ensure that individuals will almost surely learn parameters, but they do not ensure that conditional beliefs about the future given the past are converging to correct conditional beliefs. The distinction between knowledge of a stochastic process and forecasts of its subsequent sample paths is most important in learning to play a Nash equilibrium. Individuals will learn strategies, but they may not learn how their opponents are going to play. Before we illustrate this point in a game theoretic context, we examine it in some simple estimation examples.

Example 2.0: Let each $H_t = \{0, 1\}$, so that H is the set of all possible sequences of 0's and 1's. Let $\Theta = \{\mathbf{p}, \mathbf{q}\}$. The elements of Θ are sequences of probabilities: $\mathbf{p} = (\mathbf{p_1}, \mathbf{p_2}, \ldots)$ and $\mathbf{q} = (\mathbf{q_1}, \mathbf{q_2}, \ldots)$. The draws of 0's and 1's are independent over time, and either the draw at each time t gives 1 with probability p_t, or at each time t the probability of 1 is q_t. Suppose that all the p_t's and q_t's are uniformly bounded away from 0 and 1. Let us also suppose that $\mathbf{p}$ is the "true" distribution of the process. Suppose the decision-maker's job is to estimate θ. A necessary and sufficient condition for Bayes estimates of θ to be consistent is that the sum

$$\sum_{t=1}^{\infty} \left(p_t \log\left(\frac{p_t}{q_t}\right) + (1 - p_t) \log\left(\frac{1 - p_t}{1 - q_t}\right) \right)$$

diverge. On the other hand, suppose the decision-maker's job is to predict, at each time t, the probability that $h_{t+1} = 1$. The optimal prediction is p_{t+1} or q_{t+1}, depending on whether θ is $\mathbf{p}$ or $\mathbf{q}$. If the condition for consistency is satisfied, then posterior beliefs will converge to point mass on $\mathbf{p}$ ($\mathbf{q}$) $\mathbf{p}$-almost surely ($\mathbf{q}$-almost surely), and so the prediction at time t will almost surely converge to p_{t+1} (q_{t+1}). If the consistency condition fails, then $p_{t+1} - q_{t+1}$ is converging to 0, so again the prediction will almost surely converge to the correct conditional distribution, even though the decision-maker never learns θ.

In the previous example it was possible to forecast without learning the parameter. In the next two examples, parameter estimation is consistent but forecasting becomes hard. The first is similar to an example found in Kalai and Lehrer (1991).

Example 2.0: Let each $H_t = \{0, 1\}$, so that H is the set of all sequences of 0's and 1's. Let Θ index the set of all point masses on the elements of H. In other words, $\Theta = H$, and μ_θ is a point mass on the sequence θ. Let p be a number between 0 and 1. Let the prior distribution ν_p be that distribution on H which is derived from independent and identically distributed draws from each H_t which assign probability p to 0 at each date t.

Obviously the distributions ν_θ on H are all singular with respect to each other. We conclude from Theorem 2.2 that for almost all θ in the support of ν_p, posterior beliefs converge almost surely to δ_θ. But the conditional distribution of $(x_{T+1,}x_{T+2},...)$ given $(h_1,...,h_T)$ never changes with respect to T. It is always that derived from i.i.d. draws of 0's and 1's which assign probability p to 0.

In this example, learning about θ does take place, but after time T all the observer has learned about θ is its first T components. And given his beliefs, he can infer nothing about the behavior of subsequent components of θ from those he already knows. As contrived as this example may seem, this is exactly the root of the problem of convergence to Nash-like play in infinitely repeated games.

Here is another example, which, in Section 3, we will put in a game-theoretic setting. Here learning occurs, but forecasting becomes increasingly difficult over time because the sensitivity of the forecast to the parameter grows at a rate faster than that at which learning occurs.

Example 2.0: A deterministic system on the unit interval I evolves according to the "tent map" dynamic:

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1/2; \\ 2-2x & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

The initial position x_0 of the process is unknown. A Bayesian decision-maker will estimate the location of x_{T+1} given information on the sequence $x_0,...,x_T$. He will be told at the beginning of stage $T+1$ in which half of the unit interval x_T is to be found: The upper interval $U=(1/2,1]$, or the lower $D=[0,1/2]$. Thus each $H_t=\{D,U\}$. His information begins at stage 0. At the beginning of stage $T+1$, before x_{T+1} is realized, the decision-maker is asked to guess its coming location. In other words, after observing H^T he must forecast h_{T+1}.

The observations available at the beginning of date $T+1$ describe an interval of width 2^{-T}, and so Bayes estimates of x_0 are consistent. Let prior beliefs have density with respect to Lebesgue measure given by $\phi_0(x)$, with c.d.f. $\Phi_0(x)$. For Φ_0-almost all initial positions x_0, the posterior predicted distribution of x_{T+1} given the observations available at the beginning of date $T+1$ converges to the uniform distribution on the unit interval, and the probability that $h_{T+1}=D$ given previous history converges almost surely to 1/2 (see Blume and Easley (1993) for a proof). The decision-maker has ever-increasing knowledge of x_0, but the residual uncertainty is so magnified by the chaotic dynamics that predicting the location of the forthcoming state is increasingly difficult.

When is prediction possible? Given the prior predicted probability distribution on sample paths, μ_ν, let $\mu^T(\cdot \mid h_T)$ denote the conditional probability of the future given the past — a probability distribution on $\prod_{t=T+1}^{\infty} H_t$.

Theorem 2.3: Let $C_\nu = \{\theta : \mu_\theta \ll \mu_\nu\}$. For all $\theta \in C_\nu$, μ_θ^T and μ_ν^T converge together in variation norm for μ_θ-almost all sample paths.

The requirement of absolute continuity is very strong, since it is on the entire space of paths and not on any finite partial histories. This Theorem excludes, for instance, the case of i.i.d. coin flips where the prior ν on the parameter is absolutely continuous with respect to Lebesgue measure. An important case where the Theorem does apply is the case where ν has finite suport. We emphasize that this absolute continuity condition is a *sufficient/* condition. It is easy to construct examples where the convergence of posterior predictive distributions is assured yet the condition of the Theorem are violated. This Theorem is an immediate consequence of the main theorem in Blackwell and Dubins (1962), which itself is a consequence of the Radon-Nikodym and Martingale Convergence Theorems.

2.2. Bayesian Decision Problems

Now that we have the basic results on learning from an exogenous data process we consider environments where the data process is partially under the control of a decision-maker. To illustrate the common elements in the literature we build a general decision model with learning and then specialize it to various problems.

At each date t the decision-maker chooses an action $x_t \in X$. After choosing x_t he observes $y_t \in Y$ which has a distribution conditional on the history prior to date t, his current action x_t and parameter θ. This distribution is described by $\eta_\theta(\cdot \mid x_t, h_1, \ldots, h_{t-1})$, where $h_\tau = (x_\tau, y_\tau)$. Finally, he receives a reward $r_t \in \mathbf{R}$ which is a function of history, the observation and the action: $r_t = u(x_t, y_t, h_1, \ldots, h_{t-1})$.

In this framework the set of observations possible at time t is $H_t = X \times Y$. As before the set of sample histories is $H = \prod_{t=1}^{\infty} H_t$ and the set of partial histories to date t is $H^t = H_1 \times \ldots \times H_t$. The decision-maker's history dependent plan of action is described by a policy $\pi = (\pi_1, \pi_2, \ldots)$; i.e. a sequence of Borel-measurable functions $\pi_t : H^{t-1} \to P(X)$, mapping partial histories into probability distributions on actions.

For each policy π and parameter θ the probability $\mu_{\theta,\pi}$ on histories describing the data process is given by the composition of the decision-maker's policy and the distribution η_θ of observations. The decision-maker is assumed to have sufficient prior knowledge to calculate $\mu_{\theta,\pi}$ for each (θ, π). In particular, he knows the map $\theta \mapsto \eta_\theta$. Formally, this is an innocuous assumption. But for results, it is not. Anything that the decision-maker is uncertain about goes in Θ and he holds beliefs in the form of a prior probability distribution on Θ. The decision-maker then, of course, knows how the system will evolve given any θ and policy π. But there is a potential problem with this assumption. Suppose the decision-maker knows the everything up to the specification of a finite, or an at most countable, number of parameters, i.e. Θ is not just a finite dimensional set, but a set with a finite number of elements. Then we can apply the theorms in Section 2.1 to obtain convergence and predictions results. Frequently, however, we do not want to assume that the decision-maker has so much prior knowledge about the environment. In many problems it is more natural to assume that the decision-maker knows the environment up to the specification of a parameter from a finite dimensional set. In this case beliefs about parameters converge, but convergence of conditional beliefs to correct conditional beliefs is problematic. The absolute continuity condition of Theorem 2.3 will not be satisfied and Example 2.3 shows how badly behaved predictions can be even in a one dimensional world. In other problems we may want to allow Θ to be an infinite dimensional space (say all probability measures on a finite dimensional set). This case is even more problematic as the exceptional sets (those of prior measure zero) that we have ignored can be large. Feldman (1990) has built on Freedman's (1965) analysis of the consistency of Bayes estimates to construct an example of a bandit problem in which, for "most" prior beliefs (the complement of a first-category set) Bayes estimates are not consistent.

The decision-maker's objective is to choose a policy π to maximize his expected discounted reward

$$E_\nu\{E_{\mu_{\theta,\pi}}\{\sum_{t=1}^{\infty}\beta^{t-1}r_t\}\},$$

where $0 \leq \beta < 1$ is the discount factor. This is now a conventional non-stationary dynamic programming problem. Ignoring, for the moment, the question of existence of an optimal policy, let us suppose that the decision-maker has selected policy π^*. (On existence see Hinderer (1970).) To address the question of rational learning in the single agent problem we can now apply the results on Bayesian learning to see that beliefs converge almost surely (with respect to the prior) and to check conditions for consistency. Alternatively, in an equilibrium setting we need to solve the decision problem for each individual, find equilibria and then apply the learning theorem. In the remainder of this section we will briefly discuss the single agent problem in order to illustrate the known results and some issues. Learning in games and market economies will be discussed in the following sections.

2.3. The Single Agent Problem

The large literature on single agent problems with learning includes such diverse problems as the classical multi-armed bandit problems, the behavior of monopolists and perfectly competitive firms in stochastic environments, and optimal stochastic growth. We will not attempt to survey this literature. Instead, we will present one problem that illustrates the basic results and points to the issues raised in the introduction about the sensitivity of the results to the presence of intertemporal links other than belief revision. We build a simple optimal growth model using the results of Easley and Kiefer (1988), Feldman and McLennan (1989) and McLennan (1987). El-Gamal and Sundaram (1991) observed in a similar model that including capital as an intertemporal link would simplify the analysis of asymptotic behavior of the system. Nyarko (1987) makes a similar observation in an optimal control problem.

Let $Y \subset \mathbf{R}_+$ denote the set of potential outputs. In each period the decision-maker observes the previous output $y \in Y$, chooses the fraction $\alpha \in [0,1]$ of output to be consumed and a labor input $\ell \in L$, a compact subset of $\mathbf{R}$. The reward is then the utility from consumption and leisure,

$$U(\alpha, y, \ell) = u(\alpha y) + v(\ell),$$

which is inceasing in consumption, αy, and decreasing in ℓ. We assume that $u(\,\cdot\,)$ and $v(\,\cdot\,)$ are continuous and bounded.

That output not consumed, $(1-\alpha)y$, and labor, ℓ, are used to produce new output through a stochastic and partially unknown technology. The density of new output $\tilde{y}$ given investment $(1-\alpha)y$, labor ℓ and unknown parameter $\theta \in \Theta$ is $f(\tilde{y} \,|\, (1-\alpha)y, \ell, \theta)$. We assume that $\Theta = \{\theta_1, \theta_2\}$, that f is continuous in all variables, and that for any $(1-\alpha)y$, ℓ, and θ, the support of f is all of Y.

The decision-maker does not know the value of θ, instead he begins with prior beliefs ν_0 on Θ, and learns over time. We exclude the degenerate cases where all prior mass is concentrated on one parameter value.

This model is richer than the usual "learning model" in that two dynamical forces are at work: Belief revision and capital accumulation. Before studying the model in full generality, we consider the special case in which labor is the only productive input:

$$f(\,\cdot\,|\, y, \ell, \theta) \equiv f(\,\cdot\,|\, y', \ell, \theta) \qquad \text{for all} \quad y, y' \in Y.$$

We will simply write $f(\,\cdot\,|\,\ell, \theta)$ for the production function. In this case the agent should consume all output in each period, and so the optimal consumption rate is $\alpha = 1$. Despite its triviality, think of this problem as an example of a Bayesian decision problem like that described in the previous section, with parameter space Θ, action space $X = L$, observation space Y and nonrandom reward $u(y) + v(\ell)$. Now the only connection across periods is through the decision-maker's beliefs, ν_T. He thus solves a dynamic programming problem with state space $P(\Theta)$, the set of probability distributions on Θ.

Although the interpretation is different, this problem is analogous to the monopolist example studied in Easley and Kiefer (1988). With the standard assumptions made there we know that there is an optimal (stationary and deterministic) policy, $\pi^*(\nu)$, describing the labor choice for any prior and a convex/ value function $V(\nu)$, describing the value of the problem for any prior:

Theorem 2.4: (1) For any initial output y_0, there is a unique, continuous and convex solution $V^* : P(\theta) \to \mathbf{R}$ to the equation

$$V(\nu) = \sup_{\pi} E_{\nu}\big\{E_{\mu_{(\pi,\theta)}}\big\{\sum_{t=1}^{\infty} \delta^t \big(u(y_{t-1}) + v(\ell_t)\big)\big\}\big\},$$

and the optimal policies are all characterized as those policies π which attain the sup.

(2) There is a stationary and deterministic policy $\pi^* : P(\Theta) \to L$ which is optimal.

The optimal policy correspondence is upper hemi-continuous in beliefs. At any date the optimal action is selected to maximize the sum of reward and discounted expected value. When this sum is concave in actions the optimal action correspondence is convex valued. Normally in dynamic programming problems, one would make sufficient concavity assumptions to insure that the expectation of the value funtion is concave in the action. However, in learning problems the value function is convex in the state. [2] Thus in general there is no way to generate a convex valued action correspondence.

An application of the Bayes learning results in Section 2.1 shows that beliefs converge almost surely. The question is: Are limit beliefs consistent? The answer to this question depends upon whether "confounding policies" exist. *C*onfounding policies/ are policies which are (a) optimal for the discount rate $\delta = 0$, and (b) such that for some $\bar{\nu}$ in the domain, the parameter θ is not identified:

$$\begin{aligned} \pi(\bar{\nu}) &= \bar{\ell}, \\ f(\,\cdot\,|\,\bar{\ell}, \theta_1) &= f(\,\cdot\,|\,\bar{\ell}, \theta_2). \end{aligned}$$

The existence of confounding policies is important for consistency because potential limit beliefs ν_∞ and limit actions ℓ_∞ must satisfy two conditions: First, given limit beliefs, limit actions ℓ_∞ maximize one-period expected reward:

$$\begin{aligned} r(\ell, \nu_\infty) &= \delta \sum_\theta \nu_\infty(\theta) \int u(\tilde{y}) f(\tilde{y}\,|\,\ell, \theta) d\tilde{y} + v(\ell). \\ V^1(\nu_\infty) &= \sup_\ell r(\ell, \nu_\infty) \\ &= r(\ell_\infty, \nu_\infty). \end{aligned}$$

Second, limit beliefs must put mass only on parameter values which are consistent with the data generated by the limit actions.

In any Bayes decision problem with no confounding policies, Bayes learning is necessarily consistent. If confounding policies do exist, the consistency of Bayesian learning depends upon the discount factor. If the discount factor is 0, corresponding to a completely myopic decision-maker, and if prior beliefs are $\nu_0 = \bar{\nu}$, then the confounding policy will choose an action at time 1 such that θ is not identified, posterior beliefs will equal prior beliefs, and so forth for all time. Easley and Kiefer (1988) prove such a Theorem in a slightly different context. McLennan (1987) and Feldman and McLennan (1989) have shown that

[2] Convexity is a consequence of Blackwell's (1951) theorem on the value of information.

if all this is true at discount factor 0, it will remain true for small positive discount factors. Alternatively, when the discount factor is sufficiently near 1 and information is strictly valuable, the gain from learning is large enough to compensate for a deviation from the short run optimal quantity. Thus (according to Easley and Kiefer (1988)), Bayes learning will be consistent. In summary, we have the following collection of results:

Theorem 2.5: Suppose that θ_1 is the true parameter value, and suppose that V^1 is strictly convex. Then there is a $\delta^* < 1$ such that for $1 > \delta > \delta^*$, $\nu_\infty = \delta_{\theta_1}$ a.s. If the policy π is confounding for non-degenerate beliefs $\bar{\nu}$, then there is a $\bar{\delta} > 0$ such that, for all $0 < \delta < \bar{\delta}$ and $\nu_0 = \bar{\nu}$, $\nu_t = \bar{\nu}$ for all t and Bayes learning is not consistent.

These results show that it need not be optimal for an individual to learn to form statistically correct beliefs (from the point of view of the modeler who knows θ) even when learning is feasible. If "rational expectations" is interpreted to mean that decision-makers know θ, then it may be optimal not to learn to be rational. Alternatively, if "rational expectations" is interpreted to mean that decision-makers optimally use all available information, then any Bayesian decision-maker is, by hypothesis, rational.

This Theorem also has implications for continuity of the optimal policy. Suppose that, as in Easley and Kiefer's monopolist example, the optimal actions given $\nu(\theta_1) = 0$ and $\nu(\theta_1) = 1$ bracket a confounding action $\bar{l}$. Then for high discount factors the optimal policy cannot be continuous. If it were, there would be some prior $\bar{\nu}$ such that $\pi(\bar{\nu}) = \bar{l}$. This prior would be invariant under Bayesian revision, and the monopoloist would never learn. But for high discount rates, we know he must learn with probability 1 starting from any non-degenerate prior.

The potential for incomplete learning is greatly reduced if we reintroduce capital as a productive input — in other words, if we add another intertemporal link to our model. Now we must track both the labor choice and a savings choice, so the action space for the dynamic program is $X = L \times [0,1]$. As before, an optimal stationary, deterministic policy will exist. Now it takes the form $\pi : Y \times P(\theta) \to X$. In this case confounding policies are unlikely to exist. Now, assuming that θ_1 is the true parameter value, the requirement for π to be confounding is that the set

$$A = \{y \ : \ f\big(\cdot \,|\, \big(1 - \alpha(y,\bar{\nu})\big)y, \ell(y,\bar{\nu}), \theta_1\big) = f\big(\cdot \,|\, \big(1 - \alpha(y,\bar{\nu})\big)y, \ell(y,\bar{\nu}), \theta_2\big)\}$$

have full measure with respect to the density

$$f\big(\cdot \,|\, \big(1 - \alpha(y,\bar{\nu})\big)y, \ell(y,\bar{\nu}), \theta_1\big)$$

for all $y \in A$. In other words, for any y in A, optimal production has to land back within A with probability 1. Otherwise at each step there would be some probability of landing outside of A and learning something about the parameter. This information would move beliefs away from $\bar{\nu}$. This condition is very restrictive, and is unlikely to be met in any economic problem with intertemporal connections in addition to those through beliefs. This observation is summarized in the following Theorem:[3]

Theorem 2.6: Let θ_1 denote the true parameter value. Suppose that for any non-degenerate prior beliefs ν there is some set $A \subset Y \times L$ of actions such that:

1. There is an $\epsilon > 0$ such that, for all $y \in Y$ and $\nu \in P(\Theta)$, $\big((1-\alpha(y,\nu))y, \ell(y,\nu)\big) \in A$ with θ_1-probability at least ϵ.
2. There is a $\delta > 0$ such that for all $(z, \ell) \in A$, the relative entropy of model θ_1 with respect to θ_2 exceeds δ.

$$\int f(\tilde{y} \,|\, z, \ell, \theta_1) \log \frac{f(\tilde{y} \,|\, z, \ell, \theta_1)}{f(\tilde{y} \,|\, z, \ell, \theta_2)} \, d\tilde{y} > \delta.$$

Then Bayes learning is consistent.

This Theorem can be proven using the methods of Blume-Easley (1992). The idea of the Theorem is that infinitely often the decision-maker chooses actions which makes the two models uniformly different. The condition looks unusual, but is not that hard to check. We have constructed examples of the optimal growth problem where the condition of the Theorem can be verified without knowing anything about the optimal policies at all, just by relying on features of the stochastic production technology.

We conclude that although incomplete learning is possible, it is delicate. The other dynamic forces working on the decision-maker break her out of the learning "sink" caused by the confounding policy. In Section 5 we shall argue that in equilibrium models with heterogeneous agents, it is even easier for other dynamical forces to overwhelm the effects of learning dynamics.

Our analysis of the single-agent decision problem has not touched on predictability. This is a consequence of the stationarity of the stochastic environment. In stationary environments, consistency makes prediction possible. In the decision problems arising from game theory, the stochastic environment may be be non-stationary, and predictability emerges as a distinct separate inssue. We discuss these problems in the next section.

---

[3] A related Theorem can be found in El-Gamal and Sundaram (1991), which weakens our main hypothesis below but also requires continuity of the optimal policy function. But assuming continuity is problematic, because continuity is intimately tied up with learning. In the Easley and Kiefer (1988) analysis, continuity is established only for those discount rates low enough that one could fail to learn. When Bayes learning is consistent regardless of the prior, it is easy to see in the Easley-Kiefer problem that the optimal policy *must/* be discontinuous at that point in the domain of the policy function where the confounding policy fails to identify the two models.

3. Learning in Games

Learning issues are central to the interpretation of Nash equilibrium as a multi-person statistical decision theory. In this interpretation, each player solves a decision problem, and equilibrium expresses a consistency relationship between the actions of each player and the beliefs of his opponents; specifically, the support of other players' beliefs about any one player is contained in the set of best responses of that player to his own beliefs. Suppose however, that beliefs and actions are not initially configured in this fashion. Will the collection of players "learn" their way to a Nash equilibrium? Will the dynamics of posterior revision so adjust beliefs that this coordination property emerges in the course of play? This question is naturally posed in the context of repeated play when players know their own payoffs but not necesarily those of their opponents. Jordan (1991a,1991b), Kalai and Lehrer (1992a, 1992b) and Nyarko (1991,1992) study the convergence problem in repeated games. Kalai and Lehrer provide sufficient conditions for the emergence of a kind of equilibrium play in continuation games. In this section we formulate the learning problem in games and identify some assumptions that guarantee convergence to Nash equilibrium play or beliefs. We will see that Bayes rationality by itself implies very little about asymptotic play of repeated games. In order to derive powerful conclusions from Bayes rationality, such as convergence to Nash outcomes or convergence of beliefs to Nash-like beliefs, it is necessary to make further assumptions about the joint configuration of players' prior beliefs. These assumptions must guarantee the predictability of the future play of other players, in the sense discussed in the previous section. Not surprisingly, the further belief restrictions will involve some kind of joint absolute continuity of prior beliefs. It will be obvious that these conditions are difficult to meet. Even though they are only sufficient, and not necessary, for asymptotic convergence, they leave us very skeptical about the possibility for robust convergence to Nash-like behavior.

In focusing explicitly on rational learning we rule out a large number of papers. We neglect the *a*d hoc/ learning papers, such as Fudenberg and Kreps (1988) or Marimon, McGratten and Sargent (1989), which either search for learning procedures which will guarantee convergence to Nash equilibrium or investigate the learning implications of some given rule whose motivation comes from elsewhere. We also overlook papers which consider the role of learning as an adaptive process at work on a population of players, such as Fudenberg and Levine (1991). This mesh of epistemic and evolutionary reasoning we believe to be more promising than either the raw application of biological ideas to social processes or the Savage-Bayesian analysis which we now survey.

3.1. Bayesian Strategy Revision Processes

As in any sequential Bayesian decision problem, we need to identify a set of parameters, the parameter-conditional observation processes, actions and rewards. We consider N-player strategic form games where player n has a finite set S_n of possible actions. After actions are selected, each player observes the joint action vector $s \in S = S_1 \times \cdot \times S_N$ and receives the reward $u_n(s)$. The stage game described by $(S_n, u_n)_{n=1}^N$ will be repeated infinitely many times and player n discounts future rewards with discount factor $\beta_n \in [0, 1)$.

Most of the learning literature has focused on games with perfect monitoring and simultaneous move stage games, and we will do so here. After each stage, each player observes the choices of his opponents. At the beginning of round t each player will have observed the sequence of play through all the preceding stages of the game. Thus, the set of sample histories for player n is $H = \prod_{t=1}^{\infty} S$. The set of partial histories up to round t is $H_t = \prod_1^{t-1} S$ for $t > 1$, and H_1 is the null history. Finally, define $H^t = \prod_t^{\infty} S$ to be the "future history" beginning at date t, and let S_n^t denote the set of plays by player n at date t.

Now we will build a parameter space for the players' decision problem. Each player is completely defined by his *type/*. A player's type is a specification of his utility function, discount parameter, and beliefs about the other players' types. This notion seems to have some circularity to it, since player 1's type contains his beliefs about 2's type, which in turn contains 2's beliefs about 1's type, etc. Mertens and Zamir (1985) have shown that nonetheless the type space can be defined in a self-consistent manner. The set of possible types for player n is denoted by T_n, with generic element τ_n. The important thing to know about the type space T_n is that we can think of a type as a vector (θ, γ), where $\theta \in \Theta$ describes the utility and discount parameters, and $\gamma \in \Gamma$ describes the "belief heirarchy". (We will assume that Θ is a Polish space; throughout the remainder of the paper we will neglect to mention necessary measurability assumptions.) Thus $T_n = \Theta_n \times \Gamma_n$, where Θ_n is the set of potential utility functions and discount parameters for player n. Let $T = T_1 \times \cdots \times T_N$ denote the space of joint types.

Nyarko (following Jordan) has found it useful to distinguish several levels of prior beliefs. For Nyarko a prior belief is a probability distribution μ_n on $T \times H$. An "interim prior" is the probability distribution $\mu_n(\,\cdot\,|\,\gamma_n)$. The "prior" is constructed before players know who they are. The "interim prior" contains the beliefs held by player n when he knows who he is but before any play has occurred. We say "contains" and not "is" because $\mu_n(\,\cdot\,|\,\gamma_n)$ is a distribution over the future actions of all players, including player n. At stage 0 the marginal of this distribution on the actions of players other than n at stage 1 represents n's beliefs about how others will behave in the first round of play.

Although in game theory we are accustomed to thinking of the interim prior as the initial set of beliefs for each player's decision problem, it is useful for the learning problem to distinguish beliefs ex ante and ex post the arrival of information about type. The key to proving learning results is to tie players' decision problems together. We will see some learning results that do this through the interim prior beliefs, and others that place hypotheses on the (unconditional) priors.

Every player in the strategic situation we have just described is solving a sequential decision problem. These problems are coupled together, because the solution to player 1's problem determines what player 2 sees. The simultaneous solutions to the decision problem are described by a *B*ayesian Strategy Revision Process,/ a concept first introduced by Jordan. Our formulation differs slightly from his. If μ and ν are measures on spaces A and B, respectively, then $\mu \otimes \nu$ denotes the product measure on the product space $A \times B$.

Definition 3.1: A *B*ayesian Strategy Revision Process (BSRP)/ is a collection of probability distributions $\{\mu_n\}_{n=0}^N$ on $T \times H$ such that,

1. For $n \geq 1$ and μ_n-almost all types (θ, γ), $\mathbf{proj}_\Theta\, \mu_n(\,\cdot\,|\,\gamma) = \delta_\theta$.
2. For $n \geq 1$, μ_n-a.s., $(\gamma_n, h_t, s_{nt}) \in \Gamma_n \times H_t \times S_n^t$,
$$s_n \in \operatorname*{argmax}_{S_n^{t+1}} E_{\mu_n}\{u_n(\,\cdot\,, \tilde{s}_{-n}^{t+1}, \tilde{\theta}_n)\,|\,\gamma_n, h_t\}$$
3. For $n \geq 1$, $\mathbf{proj}_{S_{-n}^t}\, \mu_n\{\,\cdot\,|\,\gamma_n, h_t\}$ is almost surely a product.
4. For $n \geq 1$, $\mathbf{proj}_{T_n}\, \mu_0 = \mathbf{proj}_{T_n}\, \mu_n$, and for all t,
$$\mathop{\mathbf{proj}}_{S^t} \mu_0(\,\cdot\,|\,\gamma, h_t) = \otimes_{n>0} \mathop{\mathbf{proj}}_{S_n^t} \mu_n(\,\cdot\,|\,\gamma_n, h_t)$$

The probability distribution μ_0 is the actual joint distribution of types and actions. Condition 1 states that each player knows her own payoff function. Condition 2 states that each player chooses actions to maximize her expected utility given her beliefs. Condition 3 states that each player believes the actions of her opponents to be chosen independently conditional upon history and her type. Condition 3 without conditioning on γ_n would be a much stronger statement — close to saying that types are independent across players. Notice that all a BSRP requires is that players maximize with respect to their beliefs. Nothing has yet been said about the correctness of the beliefs.

3.2. The Content of Bayesian Learning

In general, requiring decision-makers to be good Bayesians imposes few constraints on strategy selection, as the following theorem shows. Let D denote the set of all distributions on $T \times H$ such that, if $\mu \in D$, then almost all conditional distributions $\mathbf{proj}_H\, \mu(\,\cdot\,|\,\gamma_n)$ are processes of players' choices which are independent across players and dates, and such that $\mathbf{proj}_{S_n^t}\, \mu(\,\cdot\,|\,\gamma_n)$ is almost surely an undominated mixed strategy in the stage game for player n of type τ_n.

Theorem 3.7: If $\nu \in D$, then there is a BSRP $(\mu_0, \mu_1, \cdots, \mu_N)$ with $\mu_0 = \nu$.

Proof of Theorem 3.7: Constructing such a BSRP is just a matter of constructing beliefs for each player n so as to make the policy $\mathbf{proj}_{S_n^t}\, \mu(\,\cdot\,|\,\theta)$ optimal. If p_n^t is the distribution of n's play at stage t, then since it is undominated there is a distribution q_n^t on the choices of the other player for which p_n^t is a best response. Let $\mu_n(\,\cdot\,|\,\theta_n)$ be the product of $p_n^t \otimes q_n^t$

over all t. This is the basic idea, but the actual construction is a bit more complicated due to the fact that one has to make everything be measurable with respect to θ_n. This can be done with the aid of a measurable selection theorem from the correspondence whose image is the set of beliefs that make p_n^t a best response for θ_n. So $\mu_n(\cdot \mid \theta_n)$ defines a transition probability. Integrating with respect to the marginal distribution of ν on Θ_n gives ν_n. □

If we replace D with the set of all un-weakly dominated strategies, the converse is true for sufficiently low discount rates.

Theorem 3.7 demonstrates that the hypothesis of Bayesian learning in games has, by itself, little content. Whatever power the Bayesian hypothesis possesses will only emerge when restrictions are placed on the nature of the Bayesian's beliefs. This power will only appear asymptotically, since the Bayesian hypothesis puts few restrictions on beliefs arising from small numbers of observations.

It is evident from Section 2 that posterior beliefs on opponents' types, and posterior beliefs on sequences of play will converge to some limit beliefs. In fact, under some mild assumptions, posterior beliefs on play histories will converge almost surely to point-mass at the true history. But this has no implications for play, as the following example shows:

Example 3.0: Consider a two-person repeated game for which, in the stage game, player 2 has two strategies A and B. Suppose player 1 correctly believes that player 2's strategy (not just actions) is a fixed sequence independent of history. Suppose the probability distribution representing prior beliefs is product measure with parameter p. Now in this case Bayes learning is consistent. Player 1 will ultimately assign probability 1 to the actual strategy employed by player 2. But at each stage he will always predict A with probability p and B with probability $1-p$.

Example 3.0 just places Example 2.0 in a game theoretic context. It shows that convergence of beliefs about strategies does not imply convergence in beliefs about strategies in continuation games.

3.3. The Conditional Harsanyi Hypothesis

Learning about strategies in a continuation game is a prediction problem rather than a consistency problem. To achieve consistency of predictions of future play by Bayesian learners, restrictions on prior beliefs must be assumed. Kalai and Lehrer's (1992a) approach to this issue uses the Blackwell-Dubins Theorem presented in Section 2. We will present the Kalai-Lehrer analysis within the framework of BSRP's in order to understand the nature of the restrictions on prior beliefs this approach requires.

The appropriate absolute continuity condition requires that for almost all types, the actual distribution of play is absolutely continuous with respect to each player's beliefs. If this is so, then the Blackwell-Dubins Theorem states that the conditional distributions on future play given histories and types must converge. This approach requires belief restrictions on interim prior beliefs. We call these restrictions the *C*onditional Harsanyi Hypothesis:/

Conditional Harsanyi Hypothesis:/ For all n, $\mathbf{proj}_H\, \mu_0(\,\cdot\,|\,\gamma) \ll \mathbf{proj}_H\, \mu_n(\,\cdot\,|\,\gamma_n)\, \mu_0$ almost surely.

The Conditional Harsanyi Hypothesis has two important consequences for beliefs. First, fix the type of player 1. The actual distribution of play for all types of player 2 must be absolutely continuous with respect to player 1's beliefs. Thus the actual play of player 2 cannot change too much with respect to 2's type. In particular, this will require that 2's play cannot vary too much with respect to his type. For instance, suppose that player 1 believes, given his type, that the frequency with which player 2 is going to play "left" converges almost surely to 1/2. Then actual play will also require this, *f*or almost all possible values of player 2's type./ The second observation is that the connection between player 1's beliefs and player 2's actions requires that player 2's beliefs be configured in certain ways. Suppose this configuration requires that the limit frequency of "up" for player 1 is 3/4. Then this requirement must be satisfied by the actual play of player 1. In other words, beliefs must initially satisfy a kind of consistency condition not too different from the consistency required by Nash equilibrium.

3.4. Belief Convergence for Myopic Players

The Kalai-Lehrer result is easiest to see in those BSRP's where the discount factor for each player is almost surely 0. The Kalai-Lehrer results state a conclusion about how the actual path of play far out in the game is almost like that of an approximate Nash equilibrium. This is complicated to state, but there are some clean conclusions to be had about the limit behavior of beliefs about future play. They converge to Nash equilibrium beliefs. (It may be the case, however, that the distribution of play does not converge to a mixed strategy Nash equilibrium profile. See Jordan (1991a,b) for a discussion of this point.) Let $M_n(\theta_n) = \{\sigma_1 \otimes \cdots \otimes \sigma_N \in P(S) \ : \ \sigma_n \quad \text{is a best response to} \quad \sigma_{-n}\}$. The Nash equilibria for the single stage game are $N(\theta) = \cap_n M_n(\theta_n)$. Let $\|\cdot\|$ denote the variation norm on the appropriate space of measures.

Theorem 3.8: Suppose that the BSRP $(\mu_n)_{n=0}^N$ satisfies the Conditional Harsanyi Hypothesis. Then

$$\mu_0\{(\gamma, h_\infty) \ : \ \|\mathop{\mathbf{proj}}_{S^t} \mu_n(\,\cdot\,|\,\gamma_n, h_t) - \mathop{\mathbf{proj}}_{S^t} \mu_0(\,\cdot\,|\,\gamma, h_t)\| \to 0\} = 1$$

and

$$\mu_0\{(\gamma, h_\infty) : \|\mathop{\mathbf{proj}}_{S^t} \mu_n(\,\cdot\,|\,\gamma_n, h_t) - N(\theta)\| \to 0 \quad \text{for all } t\} = 1.$$

Proof of Theorem 3.8: The second statement follows from the first and part 2 of the Bayesian Strategy Revision Process definition, which states that

$$\mu_0\{(\gamma, h_\infty) \;:\; \mathop{\mathbf{proj}}_{S^{t+1}} \mu_n(\,\cdot\,|\,\gamma_n, h_t) \in M_n(\theta_n) \quad \text{for all } t\} = 1.$$

The first statement is a consequence of the Conditional Harsanyi Hypothesis and the Blackwell-Dubins Theorem. □

Requirement 1 of the definition of a Bayesian Strategy Revision Process is unnecessary. Theorem 3.8 can be extended to include games with incomplete information about one's own type. We have reported an example of this in Blume and Easley (1992).

3.5. Subjective Equilibrium

When the discount factor is positive, matters are more complicated because at any decision node the entire future course of play, and not just the current play, is payoff-relevant. Again the Blackwell-Dubins Theorem will imply that limit predictions are correct. But this does not mean that each player eventually knows the other players' strategies, since information about "off-path play" may never be observed. It does imply that limit beliefs are stable in the sense that subsequent information gives no cause to revise them. Kalai and Lehrer (1992b) have introduced the notion of *s*ubjective equilibrium/ to summarize the notion of best-responding to beliefs which correctly predict the course of play. [4] Here is the definition for a finite game.

Definition 3.2: A *s*ubjective equilibrium (SE)/ is a strategy profile-prediction profile $2N$-tuple $(\sigma_n, \pi^n)_{n=1}^N$ where $\sigma_n \in P(S_n)$ is player n's strategy and $\pi^n \in P(S_{-n})$ is player n's (product) beliefs about the play of players $m \neq n$ such that:

1. Each σ_n is a best response to π^n;
2. For all n, $\sigma_1 \otimes \cdots \otimes \sigma_N = \sigma_n \times \pi^n$.

For repeated games, the definition is essentially the same, but more notation is required. Let $F_n = \{(f^1, \ldots) \;:\; f^t : H_t \to P(S_n)\}$ denote strategies for player n. Let $F = F_1 \times \cdots \times F_N$. If ν is a probability distribution on F, let $\sigma(\nu)$ denote its (Kuhn-equivalent) strategy in F. Let $\nu(f) \in P(H)$ denote the distribution on play induced by strategy profile f. Finally, define

$$u_n(\theta_n, h_\infty) = \sum_{t=1}^{\infty} \delta_n(\theta_n)^{t-1} u_n(\theta_n, s_t)$$
$$v_n(\theta_n, f) = E_{\nu(f)}\{u_n(\theta_n, \tilde{h}_\infty)\}$$

---

[4] Kalai and Lehrer origionally called this concept "private beliefs equilibrium".

Definition 3.3: A *subjective equilibrium* for a repeated game is a strategy profile-prediction profile $2N$-tuple $(f_n, g^n)_{n=1}^N$ where $f_n \in F_n$ is player n's strategy and $g^n = (f^{nm})_{m \neq n}$, $f^{nm} \in F_m$, is the Kuhn representation of player n's beliefs about the play of players $m \neq n$, such that:

1. For all n, each f_n is a best response to g^n: $v_n(\theta_n, f^n, g^n) \geq v_n(\theta_n, f'^n, g^n)$ for all $f'^n \in F_n$;
2. For all n, $\nu(f^n, g^n) = \nu(f)$.

Notice that the conditional distributions $\mathbf{proj}_{S_n^t}\, \mu_n\{\cdot \mid \gamma_n, h_t\}$ are a *plan* for player n. They do not define a *strategy* for player n in the traditional sense because the conditional expectations given unreached nodes are not well-defined. These conditional distributions can be extended to all of $T_n \times H_t$, and a collection of these extensions is a strategy. However these extensions are somewhat arbitrary, and this is the reason why convergence will be to a subjective equilibrium and not to a Nash equilibrium.

In two-person repeated normal form games with perfect monitoring (this excludes the multi-armed bandit problem), the set of subjective equilibrium outcomes and the set of Nash equilibrium outcomes coincide. This is a consequence of Kuhn's Theorem, which states that beliefs over strategies are themselves equivalent to strategies. We will state and prove this Theorem for the trivial case of finite games. It can be extended to repeated games with perfect monitoring.

Theorem 3.9: Consider a two-person game, and let $(\pi_1, \pi_2, \pi^1, \pi^2)$ be a subjective equilibrium. Then the prediction profile pair (π^1, π^2) is a Nash equilibrium, and $\nu(\pi_1, \pi_2) = \nu(\pi^1, \pi^2)$.

Proof of Theorem 3.9: Let R_n denote the collection of information sets belonging to player n that are reached in the equilibrium (π_1, π_2). Notice first that the π^n can be represented by strategies (Kuhn's Theorem), and $\pi^n | R_n = \pi_n | R_n$. Since the two strategy pairs agree on all reached information sets, $\nu(\pi_1, \pi_2) = \nu(\pi^1, \pi^2)$. Let $V(\pi, \pi^2, \theta_1)$ denote the expected return to 1 from playing $\pi \in \Pi_1$ conditional upon his type against strategy π^2. Then $V(\pi^1, \pi^2, \theta_1) = V(\pi_1, \pi^2, \theta_1)$ since π^1 and π_1 coincide on R_1. Since π_1 is a best response to π^2, so is π^1. $\square$

The conclusion of Theorem 3.9 remains true for general N-player normal form games and, more generally, multi-stage games with observable actions, when the beliefs are symmetric in the sense that any two players i and j share common beliefs about what k will play, and when each players' beliefs about the strategic choice of the other players are independent. Even when the symmetry condition fails, the independence hypothesis guarantees a related conclusion: There is a Nash equilibrium strategy profile $(\pi'_n)_{n=1}^N$ such that

$$\mu(\pi'_1, \ldots, \pi'_n) = \mu(\pi_1, \ldots, \pi_n) = \mu\big(\pi_n, (\pi^k)_{k \neq n}\big).$$

In general N-player games there may be subjective equilibria whose outcomes are not Nash equilibrium outcomes. See Blume and Easley (1993) for an example.

3.6. Convergence to Subjective Equilibria

The notion of BSRP's defined in section 3.1 is inadequate for discussing dynamic games, because conditional distributions from μ_n of future play of player n's opponents given a potential deviation by player n may not be well defined. This does not matter for myopic players because future play is payoff-irrelevant, but it does matter when discount rates factors are positive. We will use the same term (BSRP) to refer to the equilibrium concept with and without inclusion of repeated game strategies. The relevant definition should be clear from the context.

Definition 3.4: A Bayesian Strategy Revision Process (BSRP)/ is a collection of probability distributions $\{\mu_n\}_{n=0}^N$ on $T \times F \times H$ such that

0. for $n \geq 1$ and μ_n-almost all $\mathbf{proj}_H \, \mu_n(\,\cdot \mid f) = \nu(f)$,
1. for $n \geq 1$ and μ_n-almost all types (θ, γ), $\mathbf{proj}_\Theta \, \mu_n(\,\cdot \mid \gamma) = \delta_\theta$,
2. for $n \geq 1$, μ_n-a.s., $(\gamma_n, h_t, f_n) \in \Gamma_n \times H_t \times F_n$,

$$f_n \in \operatorname*{argmax}_{F_n} E_{\mu_n} \{ E_{\nu(f_n, f_{-n})} \{ \sum_{r=t+1}^{\infty} \delta_n(\tilde{\theta}_n)^r u_n(\theta_n, s_r) \} \mid \gamma_n, h_t \}$$

3. for $n \geq 1$, $\mathbf{proj}_{F_{-n}} \, \mu_n\{\,\cdot \mid \gamma_n, h_t\}$ is almost surely a product, and
4. for $n \geq 1$, $\mathbf{proj}_{T_n} \, \mu_0 = \mathbf{proj}_{T_n} \, \mu_n$, and for all t,

$$\mathop{\mathbf{proj}}_{S^t} \mu_0(\,\cdot \mid \gamma, h_t) = \otimes_{n>0} \mathop{\mathbf{proj}}_{S_n^t} \mu_n(\,\cdot \mid \gamma_n, h_t)$$

The interpretation of these conditions is exactly as before; they have just been rewritten to accomodate the present payoff-relevance of future play.

Bayesian Strategy Revision Processes and Subjective Equilibria are very different kinds of objects. We will ultimately show that under some conditions, BSRP's asymptotically "look like" subjective equilibria. We mean this in the sense that the beliefs about the future and the play in the BSRP satisfy the SE conditions. The following Lemma is an immediate consequence of the definitions.

Lemma 3.1: Suppose $(\mu_n)_{n=0}^N$ is a Bayesian Strategy Revision Process such that μ_0-almost surely,

$$\mathop{\mathbf{proj}}_{S^t} \mu_n(\,\cdot \mid \gamma_n, h_t) = \mathop{\mathbf{proj}}_{S^t} \mu_0(\,\cdot \mid \gamma, h_t). \tag{3.1}$$

Then $\big(\sigma(\mathbf{proj}_F \, \mu_n(\,\cdot \mid \gamma_n))\big)_{n=1}^N$ is a subjective equilibrium.

Bayesian Strategy Revision Processes have players maximizing given their beliefs and information, and the hypothesis of the Lemma states that each player correctly predicts the actual distribution of play.

Our version of the Kalai-Lehrer Theorem states that expectations over strategies of weak subsequential limits of BSRP's satisfying the Conditional Harsanyi Hypothesis are SE's.

Theorem 3.10: Suppose that the Bayesian Strategy Revision Process $(\mu_n)_{n=0}^N$ satisfies the Conditional Harsanyi Hypothesis. Then

$$\mu_0\big((\tau, h_\infty, f) \ : \ \|\operatorname*{\mathbf{proj}}_{H^t} \mu_n(\,\cdot\,|\,\gamma_n, h_t) - \operatorname*{\mathbf{proj}}_{H^t} \mu_0(\,\cdot\,|\,\gamma, h_t)\| \to 0\big) = 1.$$

Let $(\mu_n^*)_{n=0}^N$ denote a collection of measures such that (1) $(\mu_n^*)_{n=1}^N$ is a weak subsequential limit of the sequence

$$\big\{\big(\mu_n(\,\cdot\,|\,\gamma_n, h_t)\big)_{n=1}^N\big\}_{t=0}^\infty,$$

(2) $\mathbf{proj}_T\, \mu_0^* = \mathbf{proj}_T\, \mu_0$, and (3) for all t, $\mathbf{proj}_{S^t}\, \mu_0^*(\,\cdot\,|\,\gamma)$ are constructed from the μ_n^* as in condition 4 of the definition of a Bayesian Strategy Revision Process. Then μ_0-almost surely, $\sigma\big(\mathbf{proj}_F\, \mu_n^*(\,\cdot\,|\,\gamma_n)\big)_{n=0}^N$ is a Subjective Equilibrium.

Proof of Theorem 3.10: The first statement follows from the Blackwell-Dubins Theorem. To prove the second statement, observe that as a consequence of the first statement, any subsequential limit satisfies equation (3.1). Thus the claim will follow from Lemma 3.1 once it is shown that the limit is a Bayesian Strategy Revision Process. Conditions 1, 3 and 4 of the definition are clearly preserved under weak limits. We need to show that Condition 2 is preserved as well.

Lemma 3.2: Let $\{\mu_n^0\}_{n=0}^\infty$ be a BSRP, and let $\{(\mu_n^t)_{n=0}^N\}_{t=1}^\infty$ be a sequence of BSRPs. Let $K \subset T$ denote the set of types for which

$$\lim_{t\to\infty} \big(\operatorname*{\mathbf{proj}}_{H_\infty} \mu_n^t(\,\cdot\,|\,\gamma_n)\big)_{n=1}^N = \big(\operatorname*{\mathbf{proj}}_{H_\infty} \mu_n^0(\,\cdot\,|\,\gamma_n)\big)_{n=1}^N.$$

Then for all $\gamma \in K$ and n, $\mathbf{proj}_{H_\infty}\, \mu_n^0(\,\cdot\,|\,\gamma_n)$ satisfies condition 2.

Proof of Lemma 3.2: Condition 2 states that each player is solving a discounted dynamic programming problem: That the conditional distributions $\mathbf{proj}_{S_n}\, \mu_n(\,\cdot\,|\,\gamma_n, h_t)$ are an optimal solution to a dynamic programming problem specified in condition 2. A characterization of optimal plans is that, for all $\epsilon > 0$ there is a R such that for all $r > R$, the optimal plan gives a ϵ-optimal solution to the dynamic programming problem with horizon r. The horizon length R can be chosen with reference only to the discount factor and utility function, and independent of the state transition rule. Thus R can be chosen for each ϵ uniformly in the μ_n^t. We will show that this condition is preserved in the limit.

The sets H_n are of partial histories are finite, so weak convergence of the marginal distributions $\mathbf{proj}_{H_r}\,\mu_n^t(\,\cdot\,|\,\gamma_n)$ implies norm convergence. Thus for all $s < t$ the conditional probabilities $\mathbf{proj}_{S_{-n}^{s+1}}\,\mu_n^t(\,\cdot\,|\,\gamma_n, h_s)$ converge. Let $v_n^t(r)$ denote the optimal value for the r horizon problem for player n whose transition rule is given by the conditional distributions from μ_n^t. As a consequence of norm-continuity, $\lim_{t\to\infty} v_n^t(r) = v_n^0(r)$ for all r. Then given $\epsilon > 0$ and choose $r > R$. The value of the plan $\{\mathbf{proj}_{S_n}\,\mu_n(\,\cdot\,|\,\gamma_n, h_s)_{s<r})$ for the r-horizon problem is at least $v_n^t(r) - \epsilon$. Since the value of the plan is clearly continuous in $\mathbf{proj}_{H_r}\,\mu_n^t(\,\cdot\,|\,\gamma_n)$, and since the value of the problem is continuous in t, it follows that the plan $\{\mathbf{proj}_{S_n}\,\mu_n^0(\,\cdot\,|\,\gamma_n, h_s)_{s<r}$ is ϵ-optimal for the r-horizon problem with transitions $\mathbf{proj}_{S_{-n}^{s+1}}\,\mu_n^0(\,\cdot\,|\,\gamma_n, h_s)$ for $s < r$. Since for all $\epsilon > 0$ there is a R such that for all $r > R$ this plan is ϵ-optimal for the r-horizon problem, this plan is optimal, and so condition 2 is satisfied. This proves the Lemma and the Theorem. □

3.7. Other Learning Results

Much of what is known about rational learning comes from Kalai and Lehrer. Another important body of work on the dynamics of repeated games played by Bayesian players comes from Jordan (1991a, 1991b) and has been extended by Nyarko (1991, 1992). They replace the Conditional Harsanyi Hypothesis with a weaker assumption, the Harsanyi Hypothesis, that requires absolute continuity only of players' prior beliefs rather than almost-sure absolute continuity of type-conditional beliefs:

Harsanyi Hypothesis:/ For all n, $\mathbf{proj}_H\,\mu_0(\,\cdot\,) \ll \mathbf{proj}_H\,\mu_n(\,\cdot\,)$ μ_0 almost surely.

Again the goal is to characterize the asymptotic behavior of BSRP's. We will summarize these results for the 0-discount factor case. A sensible version of part 2 of the following Theorem is not yet known for positive discount factors.

Suppose that players' types are independently distributed. The main result is that for almost all type profiles $\gamma = (\gamma_1, \ldots, \gamma_N)$, the conditional distribution of beliefs on future play *g*iven history, but not types,/ converges weakly to a Nash equilibrium of the repeated game with type profile γ. If the type distribution is not a product, then the limit of the conditional distribution of beliefs on future play given history is a correlated equilibrium. Let $C(\theta)$ denote the set of all probability distributions on H_∞ that are distributions of play arising from the correlated equilibria of the game with characteristic parameters $\theta = (\theta_1, \ldots, \theta_N)$. Let

$$G = \{(\tau, h_\infty) \;:\; \mathbf{proj}_{S^{t+1}}\,\mu_n(\,\cdot\,|\,h_t) \to C(\theta) \quad \text{for all } n\},$$

where limit means weak-convergence limit. Let $\alpha(\,\cdot\,|\,h_t)$ denote the empirical distribution of play through date t. Let

$$F = \{(\tau, h_\infty) \in G \;:\; \mu_n(\,\cdot\,|\,h_t) - \alpha(\,\cdot\,|\,h_t) \to 0\},$$

again with weak convergence. The following Theorem is proven in Nyarko (1992).

Theorem 3.11: Suppose that the Bayesian Strategy Revision Process $(\mu_n)_{n=0}^N$ satisfies the Harsanyi Hypothesis. Then

1. $\mu_0(G) = 1$;
2. $\mu_0(F) = 1$.

If players' types are independent, correlated equilibrium can be replaced with Nash equilibrium. It is hard to interpret these results as statements about limits of players' beliefs, because players' beliefs are formed by conditioning on their type as well as the history of play. The one case where such an interpretation is possible is when types are independently distributed. In the language of BSRP's, this is the requirement that the projection onto T of the distribution μ_0 is a product. In this case the conditional distribution of future play given history and type is type-independent, and so the limiting distributions measured by the Jordan and Nyarko theorems are the belief distributions of the players.

Nonetheless, because of the second part of the Theorem these results provide an important epistemic foundation for Nash and correlated equilibrium which is distinct from the epistemic hypothesis explored by Kalai and Lehrer. Nyarko (1992) has proven that in a BSRP satisfying the Harsanyi Condition, the *e*mpirical distribution of play/ converges to the limit correlated or Nash equilibrium. Thus these equilibrium concepts are justified as descriptions of the average behavior of play emerging from the process of active learning. This does not justify Nash or correlated equilibrium as the stable limit of players actions as they jointly learn about the play of each other. Instead it justifies these equilibrium concepts as an observable feature of play even though players choices never settle down in the stronger sense described by Kalai and Lehrer.

The Harsanyi Condition required by Jordan and Nyarko is significantly weaker than the Conditional Harsanyi Hypothesis required for Kalai-Lehrer style results. It is not hard to build examples of BSRP's similar to Example 3.0 for which the Conditional Harsanyi Hypothesis fails, and yet the Harsanyi Hypothesis holds. Nonetheless, it seems intuitive that the set of BSRP's satisfying the Harsanyi Hypothesis is small. And if players' posterior beliefs over time averages are mutually singular, then the second conclusion of Theorem 3.11 must fail. Thus we are skeptical about the possibilities of finding a broad epistemic foundation for Nash and correlated equilibria.

4. Learning in Competitive Economies

The learning problem in competitive economies shares many features with the learning in games problem. In this section we formulate the problem and provide a positive, but limited, result. As in the previous section, our analysis draws heavily on the work by Jordan (1991a,1919b) and Nyarko (1991,1992). Arrow and Green (1973), Townsend (1978), Blume and Easley (1984), Bray and Kreps (1987) and Feldman (1987) all pose the same question that we pose here. All of these authors use equilibrium models with rational learning and so examine the long run implications of learning within a "grand rational expectations equilibrium" (Bray and Kreps (1987)). Kalai and Lehrer (1990) provide an analysis of learning in competitive economies which does not use conditioning on contemporaneous data and which focuses on learning about an equilibrium rather than learning within an equilibrium. Other than not conditioning on contemporaneous data, Kalai and Lehrer's analysis is, at a formal level, virtually identical to the analysis presented here. The primary difference lies in interpretation.

We consider a simple version of the dynamic economy analyzed by Radner (1972). Our economy has a sequence of incomplete markets; at each date there will be a spot market for the single physical good and a market for one period forward delivery of the good. To keep things simple, we do not consider uncertainty or differential information about asset payoffs. The market structure and endowments are fixed and known, but various preference profiles are possible. Given Radner's assumptions, our economy would have an equilibrium of plans, prices and price expectations for each specification of preferences. Each of these equilibria specifies (ignoring issues of multiplicity) a sequence of prices which Radner's consumers are assumed to perfectly forecast. Suppose, however, that consumers do not know initially know the price sequence. They would then learn about future prices by watching the evolution of past prices. This learning problem could be modeled with individuals learning directly about price sequences as in Kalai and Lehrer. We take an alternative approach and assume that consumers do not know preferences, but learn about them over time. (If they knew each other's preferences and the map from preferences to prices they could, in principle, compute the price sequence.)

Upon observing a price in any period each individual revises his beliefs about other's preferences and about future prices accordingly. As we allow individuals to condition on contemporaneous prices we immediately encounter the problem addressed by Radner in his 1979 paper on Rational Expectations Equilibria. Current prices may reveal to individuals information about the preferences of others, about others beliefs about the prefenence profile and so on; that is, current prices may reveal information about types. To infer this information, and to use it in forecasting future prices, each consumer needs a model of the relationship between types and prices. If individual's models are correct and markets clear at each date we have a sequence of REEs. We will not assume that individuals have correct models, but in order to learn they will need to put positive probability on the correct price system given knowledge of the type vector.

We consider an economy with I consumers. At each of an infinite sequence of dates indexed by t, consumer i receives a positive endowment e^i of the single physical good. The amount of the good consumed by i at date t is denoted c_t^i and his forward purchase for delivery of the good at date $t+1$ is denoted f_t^i. We assume that at date 1 individuals have no endowment of forward contracts, i.e. $f_0^i = 0$. Finally we let $p_t \in P = \Re_+^1$ be the price of the forward contract at t in terms of the numeraire consumption good.

We suppose that each consumer knows the functional form of everyone's utility function $U^i(c^i, \theta^i)$, where c^i is a consumption sequence for i and $\theta^i \in \Theta^i$ is a utility parameter for i. Each consumer knows his own parameter, but does not know other consumers' parameters. Consumer i's type, τ^i, specifies his parameter θ^i and hierarchy of beliefs about the joint parameter vector $\theta = (\theta^1, \ldots, \theta^I)$ just as in the previous section. Let T^i be the set of possible types for i and $T = \prod_{i=1}^I T^i$.

At date t consumer i will know the prices through date t, $p^t = (p_1, \ldots, p_t)$, the amount of the good that he has contracted for, f^i_{t-1}, and his own type, τ^i. Given this data he must decide how much to consume and how many forward contracts to purchase or sell. Let his date t demand correspondence be $(C^i_t(p^t, f^i_{t-1}, \tau^i), F^i_t(p^t, f^i_{t-1}, \tau^i))$. A plan $(C^i, F^i) = (C^i_t, F^i_t)_{t=1}^\infty$ for i specifies his demands at each date. Given a price system $p^\infty = (p_1, p_2, \ldots) \in P^\infty$, the set of feasible plans for consumer i is

$$\Gamma(p^\infty) = \{(C^i, F^i) : c^i_t + p_t f^i_t = e^i + f^i_{t-1}, c^i_t \geq 0, \quad \text{for each } (c^i_t, f^i_t) \in (F^i_t, C^i_t), \forall t\}.$$

Consumer i will choose a plan to maximize his expected utility. The maximization hypothesis is included in the following definition of an equilibrium process.

Let $f_t = (f^1_t, \ldots, f^I_t)$ and $c_t = (c^1_t, \ldots, c^I_t)$. The data that an outside observer could see at date t is $h_t = (p_t, f_t, c_t) \in H = P \times \Re^I \times \Re^I_+$. Let the history of the process to date t be $h^t = (h_1, \ldots, h_t) \in H^t$ and let H^∞ be the set of histories.

Definition 1: A *Bayesian Equilibrium Process* is a collection of probabilities $(\mu^i)_{i=0}^I$ on $T \times H^\infty$ such that:

1. For each i: $\quad \tau^i = \mathbf{proj_T}\, \mu^{\mathbf{i}}(\cdot|\tau^{\mathbf{i}})$.
2. For each i: $\quad \mu^i$ almost surely, $(c^i_t, f^i_t) \in (C^i_t(p^t, f^i_{t-1}, \tau^i), F^i_t(p^t, f^i_{t-1}, \tau^i))$, where

$$(C^i, F^i) \in \operatorname{argmax} E(U(c^i, \theta^i)|\tau^i)$$
$$(C^i, F^i) \in \Gamma(p^\infty)$$

3. Prices evolve according to the price system p^∞ where, for all t:
   (a) $\mathbf{proj_{P^\infty}}\, \mu^{\mathbf{0}}(\mathbf{p}^\infty|\tau) = 1$, where $\tau = (\tau^1, \ldots, \tau^I)$, and
   (b) Markets clear at each date,

$$\sum_i c^i_t - e^i = 0$$
$$\sum_i f^i_t = 0.$$

There are several elements of this definition that need comment. First, μ^i represents i's beliefs on types cross histories. Part 1 simply restates the definition that i's type gives his beliefs on the type space. Part 2 of the definition requires each consumer to maximize his expected utility using his beliefs. Finally, Part 3 specifies a true equilibrium price process. The artificial agent 0 has beliefs μ^0 which place probability one on the price system that is selected given the consumer's types. Thus, we have assumed that a sequence of equilibria exist for each possible draw of types. For standard utility functions proving that a Bayesian Equilibrium Process exists is straightforward. Just let each μ^i place probability one on the same price system and apply the analyis from Radner (1972). Proving the equilibrium exist for non-trivial beliefs would be difficult.

To demonstrate that our definition of an equilibrium process permits non-trivial beliefs and to show how learning occurs we provide an example of an equilibrium process. Consider an economy in which for each consumer i,

$$U^i(c^i,\theta^i) = \sum_{t=1}^{\infty}(\theta^i)^{t-1}\log(c^i_{t-1}),$$

where $0 < \theta^i < 1$. Define the mean of the discount factor distribution with respect to the endowment distribution to be $m = \sum_{i=1}^{I}\frac{\theta^i e^i}{e}$ where $e = \sum_{i=1}^{I} e^i$ and define the variance of the discount factor distribution with respect to the endowment distribution to be $v = \sum_{i=1}^{I}\frac{(\theta^i)^2 e^i}{e} - m^2$. We assume that the joint parameter vector is drawn from a distribution on

$$\{\theta \in (0,1)^I : \sum_{i=1}^{I}\frac{(\theta^i)^2 e^i}{e} - m^2 = v\}$$

where the value of v and the distribution are common knowledge.

We construct a Bayesian Equilibrium Process for this economy by first finding an equilibrium of plans, prices and price expectations. Let $z_t = 1+p_{t+1}+p_{t+1}p_{t+2}+p_{t+1}p_{t+2}p_{t+3}+\cdots$. Calculation shows that

$$f^i_t = \frac{\theta^i(e^i+f^i_{t-1}) - (1-\theta^i)e^i p_t z_t}{p_t}.$$

In an equilibrium these forward demands must sum to zero in each period. So we have for each t,

$$p_t z_t = \frac{\sum_i \theta^i(e^i+f^i_{t-1})}{\sum_i(1-\theta^i)e^i}.$$

This equation system, along with the initial condition $f_0 \equiv 0$, determines the equilibrium price sequence. To be specific, in period 1,

$$p_1 = m - \frac{v}{(1-m)}$$

$$z_1 = \frac{m}{(m-(v+m^2))}$$

The prices above were calculated under an assumption that each consumer knew all future prices, but as consumers do not know θ they cannot calculate these prices prior to first period trade. Notice, however, that first period futures demands depend on future prices only through z_1. Given knowledge of v, and the equation system determining equilibrium prices, p_1 reveals m. So each consumer can calculate z_1. In each period beyond the first, z_t can be inferred from p_t and knowledge of z_{t-1}. Thus, the equilibrium of plans, prices and price expectations above is also a sequence of rational expectations equilibria for this economy.

The common priors in the description of the economy and the equilibrium above define a Bayesian Equilibrium Process for the economy. This equilibrium is, even in the first period, an equilibrium of plans, prices and price expectations. More interesting would be an equilibrium in which consumers eventually, but not instantly, learn all payoff relevant information. Bray and Kreps (1987) provide, in a slightly different model, an example of this sort. They focus on an economy with no intertemporal links other than learning and show that agents' beliefs eventually converge to "correct beliefs". The results of Jordan (1982) suggest, however, that in the typical economy revelation, or near revelation, in the first period may be possible. Jordan shows (in a model that is different from the model here) that in the generic economy, with the dimension of the space of private information greater than the number of relative prices, approximately full revelation occurs in a REE. This remarkable result uses a map from private information to prices which is pathalogical. The approach to learning taken in this section assumes that individuals know this map, or at least put positive prior probability on it, and use their prior knowledge to learn other's private information. This can at best produce partial answers to the learning question.

At this point it is important to be careful about what is meant by the "learning question". At one level the question is do individuals learn each others type? The more important question is does the sequence of equilibria converge to an equilibrium with common, correct expectations for the underlying true economy? A positive answer to the first question does not insure a positive answer to the second. In both the Bray and Kreps (1987) and Blume and Easley (1984) papers a positive answer to the first question is used to produce a positive answer to the second. However, both papers are about stationary economies with no intertemporal connections other than learning. Further, Bray and Kreps exploit the continuity that their example produces and Blume and Easley look at an economy with only a finite number of types.

We first examine the question of learning types. Convergence of beliefs about types follows immediately from Theorem 2.1.

Corollary 4.1: Posterior beliefs on types $\mathbf{proj_T}\,\mu^{\mathbf{i}}(\cdot|\mathbf{p^t},\tau^{\mathbf{i}})$ converge τ^i -almost surely.

This result does not say that individuals learn τ. If the price system does not reveal τ it could not be learned, but then as the example above shows learning types only up to the equivalence classes induced by the price system is sufficient. An application of Corollary

2.1produces the desired result.

Corollary 4.2: Let A be a Borel subset of T and let P_A denote the subset of paths $p^\infty \in P^\infty$ such that p^∞ is not in $\operatorname{supp} \mathbf{proj_{P^\infty}}\, \mu^{\mathbf{i}}(\cdot|\tau)$ for any $\tau \in A$. For $\mathbf{proj_{P^\infty}}\, \mu^{\mathbf{i}}$ -almost all $p^\infty \in P_A$, the posterior probability $\mathbf{proj_T}\, \mu^{\mathbf{i}}(\mathbf{A}|\mathbf{p^t})$ converges to 0.

So from i's point of view he will learn the type vector up to the equivalence classes induced by prices.

The examples in Section 2 show that learning the parameter is not necessarily enough to insure that individuals conditional forecasts about the future given the past converge to correct conditional beliefs. We have not translated these examples into an equilibrium setting because constructing equilibria that do not reveal types in the first period is difficult. But we believe that the points made in Section 2 about the difference between learning types and learning to forecast are valid in an equilibrium setting.

To prove that consumer's conditional beliefs about future prices converge to the true conditional distribution of future prices we need some structure on initial beliefs. The following is a translation of the Conditional Harsanyi Hypothesis to our competitive framework.

Axiom 1 $\mathbf{proj_{T\times P_1}}\, \mu^{\mathbf{0}}$ almost surely $\mathbf{proj_{P^\infty}}\, \mu^{\mathbf{i}}(\cdot|\mathbf{p_1}, \tau^{\mathbf{i}}) \gg \mathbf{proj_{P^\infty}}\, \mu^{\mathbf{0}}(\cdot|\mathbf{p_1}, \tau)$, for all i.

The absolute continuity assumption above is placed on individual's beliefs about price systems at the time decisions are made. This is after they have seen their own type and the first period price. In any economy in which the first period price reveals types up to the equivalence classes induced by price systems this assumption is easy to satisfy. But in this case the assumption is almost the result. Consumers are not assumed to know what prices will occur or to perfectly forecast future prices given the first period price, but they are assumed to place strictly positive probability on the prices that will occur given the first period price.

The following result is an immediate consequence of the Blackwell and Dubins (1963) theorem on the merging of opinions. It implies that outside of a set of μ^0 measure zero consumers' beliefs converge together and converge to the true conditional distribution $\mathbf{proj_{P^\infty}}\, \mu^{\mathbf{0}}(\cdot|\tau)$. Thus the economy will eventually be approximately in an equilibrium of plans, prices and price expectations.

Corollary 4.3: Suppose that Axiom 1 holds. Let $\mu^i(\cdot|p^t, \tau^i)$ be consumer i's date t conditional probability given his observation of prices up to date t and his own type. Then

$$\mathop{\mathbf{proj}}_{\mathbf{P}} \mu^{\mathbf{i}}(\mathbf{p_{t+1}}|\mathbf{p^t}, \tau^{\mathbf{i}}) \to 1 \quad \text{almost surely } \mu^0.$$

Corollary 4.3 would be true with Axiom 1 replaced by an absolute continuity assumption on beliefs conditioned only on types and not on first period prices. But this revised axiom would be very restrictive. It requires each $\mathbf{proj_{P^\infty}}\,\mu^{\mathbf{i}}(\cdot|\tau^{\mathbf{i}})$ to have an atom. This effectively restricts the type space to a finite or countably infinite set. Unlike the axiom that we use this assumption is not met in the example above.

Kalai and Lehrer's (1990) paper on learning in equilibrium begins with beliefs directly over price systems. They show that under an absolute continuity assumption on beliefs (not conditioned on first period prices) learning leads to an equilibrium of plans, prices and price expectations. Although we begin with beliefs over parameters, these beliefs induce beliefs over price systems. Further the stronger absolute continuity assumption refered to above, when translated to beliefs over price systems, is essentially the same as Kalai and Lehrer's assumption.

5. Robustness of Equilibrium Learning

In the preceding section we examined the implications of incredibly sophisticated Bayesian learning. Individuals were assumed to place positive prior probability on the correct model and to behave exactly as Bayes rule requires in updating their beliefs. We believe that both of these assumptions are unrealistic. In this section we consider, in a simpler model than was used in the last section, what happens if we drop either of these assumptions. We perturb Bayes learning in two ways, by misspecifying the model and by employing different updating rules, and show in each case how the asymptotic posterior distributions change. Our results suggest that the ability to learn is not a robust property of learning rules.

Even if the belief adjustment dynamic fails there may be other forces at work in the economy that cause the sequence of temporary equilibria to converge to a rational expectations equilibrium. For some economies it is sufficient to have only some Bayesian traders as the non-Bayesians will be driven out of the market by the Bayesians. In other economies, however, the market may select for traders who do not even asymptotically have rational expectations. To analyze the market selection force we use a simple model general equilibrium model with uncertainty, but without differential information.

5.1. A Prototype Economy

In this section we introduce a prototype economy in which our points are easily illustrated. Our economy contains one good, which can either be invested or consumed. At the end of period $t-1$, investors choose a portfolio of investments that pay out one period hence contingent upon the occurrence of some event. At the beginning of date t, the state of nature is realized and the assets pay out. Investors then decide how much to consume, and how to invest that which is not consumed in assets which pay out at the beginning of date $t+1$.

Investment payouts are state-contingent, but investors do not know which state will be realized at the time of their investment. To present the learning model in its simplest form, we also suppose that no investor has inside information about the state to be realized. The most that any investor can hope to know is the stochastic process generating the states. Investors do not even know this, but they can learn it. Here $r$ational expectations/ means that investors know the true distribution of future states.

Now we turn to formalities. Time is discrete and indexed by t. There are 2 states of the world, indexed by $s \in \{0, 1\}$, one of which will occur at each date. States follow an i.i.d. process with probability $1 > q > 0$ of state 0 being realized at any date. Let X_t denote the random variable whose value is the realization of the process at date t. Let $(X, \mathcal{F}, \mu)$ denote the measurable space of values of the process ($X = \prod_{t=1}^{\infty} X_t$, $\mathcal{F}$ is the product σ-field and μ is the product measure induced by q), and let $\mathcal{F}_t$ denote the σ-field of events measurable through time t.

At each date there is one unit of each of 2 assets available. If state s occurs at date t then asset s pays off \$1 and all other assets have a zero payoff. So total wealth in the economy at date t will be \$1 regardless of which state occurs. This wealth will be distributed among the traders proportionately according to the share of the successful asset each trader owns.

Let w_{t-1}^i denote trader i's wealth net of date $t-1$ consumption, and let α_{st}^i denote the fraction of that wealth he invests in asset 0 that pays off at the beginning of date t. We assume that all wealth is invested either in asset 0 or in asset 1. (Any money not invested disappears.) The price of asset st (which pays out in sate s at date t) is denoted by ρ_{st}. So trader i owns $\alpha_{st}^i w_{t-1}^i / \rho_{st}$ shares of asset st at the end of date $t-1$. Thus, his investment income at date t is $(\alpha_{st}^i w_{t-1}^i / \rho_{st})$ if state s occurs.

After realizing his investment income, trader i consumes fraction $1 - \delta_{st}^i$ and saves fraction δ_{st}^i to invest in assets paying out at date $t+1$. So if state s occurs at date t, trader i's wealth at the end of date t will be $\delta_{st}^i(\alpha_{st}^i w_{t-1}^i / \rho_{st})$. We refer to $\{\alpha_{0t}^i\}_{t=1}^{\infty}$ as trader i's portfolio rule and the pair $\{\alpha_{0t}^i, \delta_t^i\}_{t=1}^{\infty}$ as trader i's investment rule.

Given the trader's wealths and portfolio rules the asset prices must satisfy:

$$\sum_{i=1}^{I} \frac{\alpha_{0t}^i w_{t-1}^i}{\rho_{0t}} = 1,$$
$$\sum_{i=1}^{I} \frac{\alpha_{1t}^i w_{t-1}^i}{\rho_{1t}} = 1.$$

Let $p_{st} = \rho_{st}/w_{t-1}$ be a normalized asset price where w_{t-1} is the market wealth at the beginning of date t. Let $r_{t-1}^i = w_{t-1}^i / w_{t-1}$ be trader i's wealth share. Then in equilibrium:

$$p_{st} = \sum_{i=1}^{I} \alpha_{st}^i r_{t-1}^i \tag{5.2}$$

and

$$p_{0t} + p_{1t} = 1.$$

The variables α^i_{st} and δ^i_{st} describe the demand for assets and for the consumption good at each date-event pair. Typically they would depend on current market prices and wealth levels, and perhaps on previously observed information as well. We will derive them from particular preferences.

We suppose that our investors are dynamic programmers. The reward function for current period consumption is $u(c) = \log c$, and investor i has discount factor δ^i. Investors are learners. Each investor believes that the true probability of state 0 is either θ^i_a or θ^i_b. Her prior belief that model a holds is ν^i_0. After each trading date, the trader observes the state realized at the beginning of that day. Her posterior beliefs after t such observations is ν^i_t. The predicted probability of state 0 is then $q^i_t = \nu^i_t\theta^i_a + (1-\nu^i_t)\theta^i_b$. Solving the dynamic program as in Section 2.2, we find that

$$\begin{aligned} \alpha^i_{0t} &= q^i_t, \\ \delta^i_{st} &= \delta^i. \end{aligned} \tag{5.3}$$

The "problem" of rational expectations takes a very simple form here. There is no information asymmetry which can lead to market failure if expectations fail to be rational. There are no information flows from informed to uninformed traders. Nonetheless, with respect to the dynamics of expectation formation, there is no difference between this example and the learning dynamics in the previous section.

We use this example to study three aspects of equilibrium behavior with learning dynamics: We investigate the effects of misspecified models, we investigate the robustness of convergence with respect to the updating rule, and we investigate the interaction between learning dynamics and the dynamics of wealth distribution.

5.2. Misspecified Models

The first question we take up is the workings of learning dynamics when every model being considered by the traders is misspecified. This issue does not appear to be very interesting in the example economy just described. It is of greater concern when there exist information asymmetries, and when traders are learning some structural features of the economy which are required to infer from prices to contemporaneous information. The importance of the model misspecification issue in this context arises from the fact that the structural model of the economy depends upon, among other things, traders' beliefs about the unknown structural parameter. If traders recognize this dependence, and their structural model of the economy takes this into account, Bayesian learning can lead to rational expectations. This is the point of Blume-Easley (1984), Bray-Kreps (1987) and Feldman (1987). If, on the other hand, traders do not recognize this dependence, then models that are correct when traders fully believe them are incorrectly specified when traders give some weight to other hypotheses. Blume-Easley (1982) show that as a consequence of this incorrect specification, beliefs may not converge to rational expectations, and the dynamics of temporary equilibrium driven by learning lead to arbitrary outcomes, including convergence to incorrect beliefs and non-convergence.

Model misspecification in our prototype example is easy to model. Suppose that for no i is it the case that θ_a^i or θ_b^i equals q. How does the economy behave in the long run? The first question is, what are limit beliefs? It seems to be well-known (and easy to prove in the i.i.d. case) that limit posterior beliefs will put mass 1 on the model which is closest to the "true" model in the sense of relative entropy. For two probability measures q and θ on $\{0, 1\}$, the relative entropy of q with respect to θ is

$$I_q(\theta) = q \log \frac{q}{\theta} + (1-q) \log \frac{1-q}{1-\theta}.$$

The expression $I_q(\theta)$ is non-negative, and 0 only when $\theta = q$. However, relative entropy is not symmetric and fails to satisfy the triangle inequality, so it is not even a pseudometric on its domain. If $I_q(\theta_a^i) < I_q(\theta_b^i)$, then trader i's posterior distributions converge to point mass on θ_a^i. If $I_q(\theta_a^i) > I_q(\theta_b^i)$, then posteriors converge to point mass at θ_b^i. Finally, if $I_q(\theta_a^i) = I_q(\theta_b^i)$, the log of posterior odds is a random walk.

Now we can describe the limit behavior of equilibrium prices. Suppose that there is one trader, say trader 1, who puts some prior weight on a model, say θ_a^1, which is closer to q (in terms of relative entropy) than any other model receiving positive weight from any other trader. Then trader 1's posterior beliefs will converge to point mass at θ_a^1, and q_t^1 converges to θ_a^1. In this case one can show (using the techniques in Blume-Easley (1992)) that the wealth share of trader 1 converges to 1, and market prices p_t converge to θ_a^1. In no sense are the assets correctly priced, but assets are priced according to the best beliefs in the market.

Theorem 5.12: If $I_q(\theta_a^1) < I_q(\theta_z^i)$, for $z = a, b$ and all $i > 1$, and $z = b$ and $i = 1$ then $p_t \to \theta_a^1$ almost surely.

Next suppose that $I_q(\theta_a^2) = I_q(\theta_a^1)$, and these two models are closer to q than all other models. Then the wealth share of all traders but traders 1 and 2 fall to 0. The wealth shares of traders 1 and 2 oscillate between 0 and 1 (with limsups and liminfs of 1 and 0, respectively), and market prices have two accumulation points, θ_a^1 and θ_a^2.

Theorem 5.13: If $I_q(\theta_a^1) = I_q(\theta_a^2) < I_q(\theta_z^i)$, for $z = a, b$ and all $i > 2$, and $z = b$ and $i = 1, 2$ then the almost sure limit points of the price sequence p_t are precisely θ_a^1 and θ_a^2.

This result also follows from the analysis in Blume-Easley (1992), and shows that, again, assets can be no better priced than by the best beliefs in the market.

These results are straightforward because the notion of "best model", meaning closest in relative entropy to the true model, is exogenously fixed. In the economies of Blume-Easley (1982) this is no longer (necessarily) the case. In these economies traders are trying to learn about the equilibrium price correspondence, and the misspecification results from the fact that traders do not take account of the effects of their (and others') beliefs on the correspondence. Now one can imagine more complicated dynamics. A trader may start off with a "best model", but as she becomes wealthier and as her beliefs put more and more weight on the best model, the equilibrium price correspondence may shift in such a way that the original "best model" is no longer best.

5.3. Robustness of Bayes Updating

Bayesian updating is a very delicate matter. The manner in which current observations and prior beliefs are combined is balanced so that, on the one hand, beliefs converge, and, on the other hand, limit beliefs are correct whenever it is possible to distinguish the truth in the data. If decision-makers put too much weight on their prior beliefs, or too much weight on the data, one or the other of these properties is lost. We will demonstrate this for the case of learning q, and explore its implications for the long run behavior of prices in the prototype economy.

Consider a Bayesian decision-maker who is undecided between two models, θ_a and θ_b. Now we suppose that $\theta_a = q$, so a Bayesian decision-maker's posterior beliefs would converge almost surely to point mass at $\theta_a = q$. But now we are going to suppose that our decision-maker is not a true Bayesian. The log of the likelihood ratio for the two models is:

$$L(X_t) = (1 - X_t)\log\left(\frac{\theta_a}{\theta_b}\right) + X_t \log\left(\frac{1-\theta_a}{1-\theta_b}\right).$$

A Bayesian decision-maker would update posterior beliefs according to the rule:

$$\log\frac{P_t(\theta_a)}{P_t(\theta_b)} = L(X_t) + \log\frac{P_{t-1}(\theta_a)}{P_{t-1}(\theta_b)},$$

where P_t is the posterior belief distribution after t observations.

We suppose instead that the decision-maker updates beliefs according to the following rule:

$$\begin{aligned}\log\frac{P_t(\theta_a)}{P_t(\theta_b)} &= (1+\lambda)L(X_t) + (1-\lambda)\log\frac{P_{t-1}(\theta_a)}{P_{t-1}(\theta_b)}\\ &= (1+\lambda)\sum_{s=0}^{t-1}(1-\lambda)^s L(X_{t-s}) + (1-\lambda)^t \log\frac{P_0(\theta_a)}{P_0(\theta_b)}.\end{aligned}$$

The case of $\lambda = 0$ corresponds to Bayesian updating. If $\lambda > 0$, then the decisionmaker puts too much weight on the data, while if $\lambda < 0$, the decisionmaker puts too much emphasis on her beliefs. A negative value for λ is not really sensible, but we include it for completeness.[5]

If $\lambda > 0$ the effect of the prior beliefs vanishes, as it does in the case of Bayesian revision. Let $Z_t = \sum_{s=0}^{t-1}(1-\lambda)^s L(X_{t-s})$. The process Z_t satisfies the difference equation $Z_{t+1} = (1-\lambda)Z_t + L(X_{t+1})$. It should be clear that the random variables Z_t are uniformly bounded by $\lambda^{-1}\log(\theta_a/\theta_b)$ and $\lambda^{-1}\log\big((1-\theta_a)/(1-\theta_b)\big)$, and that they do not converge. Thus $\log\big(P_t(\theta_a)/P_t(\theta_b)\big)$ does not converge, and is uniformly bounded away from $-\infty$ and $+\infty$.

If $\lambda < 0$, take $a = 1-\lambda$ and consider:

$$\frac{1}{\sum_{s=0}^{t} a^s}\log\frac{P_t(\theta_a)}{P_t(\theta_b)} = (1+\lambda)\frac{\sum_{s=0}^{t-1} a^s L(X_{t-s})}{\sum_{s=0}^{t-1} a^s} + \left(\frac{a^t}{\sum_{s=0}^{t-1} a^s}\right)\log\frac{P_0(\theta_a)}{P_0(\theta_b)}.$$

The last term on the right converges to $-\lambda\log P_0(\theta_a)/P_0(\theta_b)$. The first term converges to:

$$-\frac{(1+\lambda)\lambda}{(1-\lambda)}\sum_{t=0}^{\infty}\left(\frac{1}{1-\lambda}\right)^t L(X_{1+t}).$$

It follows from the Martingale Convergence Theorem that this "discounted sum" converges, and so the right hand side converges to some limit random variable. Clearly for "most" prior beliefs, the right hand limit will almost surely not be 0, so posterior beliefs must converge to 0 or 1 (since the denominator on the left hand side is diverging). But in this case the limit beliefs need not be correct. For instance, suppose that prior beliefs assign equal probability to θ_a and θ_b so that the log of the prior odds ratio is 0. It is easy to see that the limit rhs random variable exceeds 0 with positive probability, and that with positive probability the limit rhs random variable is exceeded by 0. Thus under θ_b the probability that posterior beliefs on θ_b go to 1 and the probability that posterior beliefs on θ_b go to 0 are both positive. Alternatively, if prior beliefs on θ_a are sufficiently large (small), then limit posterior beliefs assign probability 1 (0) to θ_a regardless of the data.

In summary, we have the following Theorem:

Theorem 5.14: If decision-makers put too much weight on the data ($\lambda > 0$), then posterior beliefs do not converge, and predicted distributions are convex combinations of the form $\alpha\theta_a + (1-\alpha_a)\theta_b$, where α is uniformly bounded away from 0 and 1.

If decision-makers put too much weight on their prior beliefs ($\lambda < 0$), then almost surely posterior beliefs converge to point mass at $\theta_a = q$ or θ_b. If the prior odds ratio

---

[5] Suppose, for example, that the models are equally likely given the data. Then if $\lambda < 0$ the posterior beliefs on θ_a go to one if $P_0(\theta_a) > P_0(\theta_b)$ and to zero in the opposite case.

is sufficiently near 1, then the limit probability of each point mass is positive. If the prior odds ratio is sufficiently different than 1, then limit posterior beliefs will put probability 1 on that model which was initially regarded as more likely.

We conclude that, when beliefs and data are incorrectly balanced in the updating formula for posterior odds, the posterior revision process will be inconsistent — correct beliefs will fail to almost-surely emerge.

Now we turn to the question of long run prices. Let us assume that for all traders, $\theta_a^i = q$ and $\theta_b^i = \theta_b > q$. Thus all traders consider the same models. Suppose first that traders put too much weight on the data ($\lambda > 0$). Then each trader's predicted distributions q_t^i will not converge, but will bounce around on some closed interval contained in (q, θ_b). As prices are a wealth share weighted average of beliefs we can conclude that, in the limit, prices move in that same interval. Notice that prices do not converge, and that prices are biased — the market odds ratio is always higher than the true odds ratio.

Theorem 5.15: If traders put too much weight on the data, then market prices do not converge. If q is an extreme point of the set of models considered by the traders, then the market price will be systematically biased (too high or too low, depending on the position of q).

When traders put too much weight on their prior beliefs, a variety of things can happen. Suppose that trader 1 assigns sufficiently high prior probability to the correct model. Then her posterior beliefs will converge to point mass on the correct model, and her predicted distribution q_t^1 converges to q. The wealth share of all traders with beliefs like hers converges to 1, and the equilibrium price converges to q. Suppose, on the other hand, that all traders place too much prior weight on the false model. Then all beliefs converge to the false model, and the market price converges to θ_b. Finally, if the prior odds of all traders are sufficiently near 1, then the updating dynamics is (with positive probability) driven by the data (with earliest observations getting the most weight). In this case, posterior beliefs converge either to point mass at q or at θ_b, predicted distributions converge either to q or to θ_b, and each happens with positive probability. However, all traders see the same information, and so all posterior beliefs move together. It is not the case that some traders will ultimately predict q and others will simultaneously predict θ_b. Thus market prices will converge either to q or to θ_b, each with positive probability.

Theorem 5.16: If traders put too much weight on their prior beliefs, then, depending upon what the prior beliefs are, market prices will converge either to q with probability 1, to θ_b with probability 1, or to each with positive probability.

When traders put too much weight on their prior beliefs, convergence to "correct" prices in the limit, when it occurs is an accident of prior specification or fortuitous data-gathering.

5.4. Learning Dynamics and Wealth Accumulation

Throughout most of the literature on learning in GE models, the dynamics of expectations adjustment provides the only link between temporary equilibria at different dates. In this section we provide examples to demonstrate the variety of ways in which learning can interact with other intertemporal connections to determine the long run behavior of equilibrium prices. In our prototype economy, the additional intertemporal connection comes from the dynamics of wealth share adjustment. Over time, some traders prosper and others suffer. The prosperous traders come to dominate the market, and the equilibrium price reflects their beliefs. This much is evident from equation (5.2). One can imagine two possible scenarios: First, learning is reinforced by wealth dynamics. Those traders with more accurate beliefs are rewarded by the market and come to dominate it. If some traders are true Bayesians, then in the long run their beliefs are accurate, they will dominate the market, and the asset will be priced correctly. Another possible scenario is that differences in decision rules more than compensate for differences in learning rules, and so rational learners may be driven from the market. In Blume-Easley (1992) we give examples of both phenomena, and we will quickly summarize these examples here.

First we will describe a situation where the dynamics of Bayesian learning and the dynamics of wealth adjustment complement one another. Suppose that all traders have log reward functions and identical discount factors, and suppose that some subset of traders consists of Bayesian learners who put positive prior probability to the model q. Then those traders' predicted distributions will almost surely converge to q, their collective wealth share will converge to 1, and market prices will converge to q. (This result is proven in Blume-Easley (1992).)

Theorem 5.17: If all traders employ identical decision rules derived from logarithmic preferences (with identical discount factors), and if some traders are Bayesian learners who put positive probability on the correct model, then assets are correctly priced in the long run.

Theorem 5.17 is surprisingly delicate. If traders use different decision rules, or if traders are heterogeneous in an asymmetric way, then the conclusion no longer holds.

Theorem 5.18: Suppose some traders have logarithmic preferences with discount rate δ and believe with probability 1 that the correct model is q. The remaining traders have logarithmic preferences, are certain that the true model is r, and have discount rate γ. If

$$I_q(r) - \log \gamma < -\log \delta,$$

then the market price process will converge almost surely to r.

This Theorem, which is a consequence of results in Blume-Easley (1992), shows that the higher savings rates of the incorrectly informed traders overwhelms the better information of the correctly informed traders. Consequently, in order to ensure that market prices converge to q, we would have to assume that information is uncorrelated with rates of time preferences. This certainly would not be true if information-gathering was costly.

If traders' reward functions are not logarithmic, then it is possible again that the market would favor traders with incorrect beliefs over those with correct beliefs. We have shown that the market selects over decisions — not beliefs, and that the market will select for those traders whose decisions α^i_{0t} are, on average, nearest to q in the sense of relative entropy. Thus the market will tend to price assets correctly, but may do so by selecting for people with incorrect beliefs because those beliefs, when operated on by the decision rule, give better decisions according to the relative entropy criterion than do those beliefs which are more accurate. In this case the market prices assets correctly, but for reasons having nothing to do with rational expectations.

6. Conclusion

In both single-agent and multi-agent sequential decision problems, the outcome of the analysis is driven by agents' expectations about the internal decision environment and exogenous payoff-relevant events. "Learning" is a device which delimits, at least asymptotically, the set of possible or realizable expectations. In single-agent decision problems, the possibilities are delimited by the choice of a prior distribution representing initial beliefs of the agent. In a single-agent decision problem, the requirements for "rational learning" amount to saying that the true parameter value is in the support of the agent's prior beliefs, and that, for every parameter value, the agent knows the likelihood function that would obtain if that parameter value was controlling the evolution of the observations. Even with these assumptions, the asymptotic outcome of the learning process may be incomplete learning, but consistency often occurs.

In multi-agent decision problems, the situation is more complicated. "Rational expectations" represents an attempt to pin down expectations by assuming that the expectations are consistent with the true structure of the decision environment. In some economic equilibrium models this is insufficiently restrictive — many rational expectations equilibria exist. Even when the equilibrium set is small, rational expectations still pose a problem. The knowledge requirements are so great that it is implausible to assume that decisionmakers just happen to be endowed with correct expectations. Hence one naturally asks if decisionmakers can learn correct expectations.

In non-cooperative incomplete information repeated game models, the Bayes-Nash equilibrium concept has embedded in it the idea that players learn over the course of play. Here Jordan (1991a, 1991b) asks if the result of this learning activity pins down beliefs as the game is repeated. As is the case in economic equilibrium models, the rationality requirements of Bayes-Nash equilibrium are heavy. Responding to this, the focus of Nyarko's and Kalai and Lehrer's research has been to ask if the rationality requirement that players *k*now/ each other's strategic choice can be relaxed so that players can *l*earn/ to play equilibrium strategies.

The crucial issue in rational learning in multiagent settings has to do with identifying the proper parameter set. Consider an equilibrium model in which a payoff-relevant signal is observed only by some traders. Suppose that the uninformed traders do not know the signal-price relationship, and will try to learn it by looking at market prices and the signal at the end of each market period. Rational learning requires that traders place positive prior probability on the true model for the *e*ntire stochastic process,/ and not just for what would happen after beliefs converged. Suppose traders know the distribution of endowments, utilities and priors on the signal process. Then the likelihood functions are, in principle, knowable. Suppose, however, that no agent knows other agents' priors. Since the evolution of the economy depends both on the original parameter value and the prior beliefs, prior beliefs on the signal process have to be added to the parameter space. Now agents must have priors on this expanded parameter space — priors on parameters cross signal-process priors. And so forth. The natural parameter space is very large. Nyarko (1991) has carried out this construction for some simple game problems. But with a large parameter space, Bayesian learning will typically fail to yield corect conditional beliefs or even to be consistent.[6] If we, the modelers, assume a simple parameterization of the choice environment, we are closing our models in the *a*d hoc/ fashion that rational learning was introduced to avoid. If we assume the natural complex parameterization, all we know is that the Bayesian believes that his beliefs will converge somewhere with probability one.

Throughout the paper we have argued that perhaps too much is being asked of learning dynamics. In economic equilibrium analysis, learning is usually studied in models where the dynamics of belief revision provide the only intertemporal link. But the results of Section 5 suggest that when other intertemporal connections are present, learning will interact with these other forces in a complicated way, and may even be irrelevant to the asymptotic behavior of the model. Similarly in the single-agent decision problem, the results of Nyarko (1987) and the growth model discussed in Section 2 suggest that the failure of learning a parameter of the state-transition equation (or conditional probability) due to (asymptotic) underidentification of the parameter, such as in Easley-Kiefer (1988) and Feldman-McLennan (1989), is largely a feature of models in which learning is the only intertemporal connection.

Arrow K. and J. Green (1973), "Notes on Expectations Equilibria in Bayesian Settings," unpublished,Stanford University.

Barron, Andrew (1989), "The Consistency of Bayes Estimators of Probability Density Functions," unpublished, University of Illinois.

Blackwell, D. (1951), "The Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley, University of

[6] See Diaconis and Freedman (1986) for a discussion of the consistency problem, and Feldman (1990) for an application to a decision problem.

California Press.

Blackwell, D. and L. Dubins (1962), “Merging of Opinions with Increasing Information,” *Annals of Mathematical Statistics,* 33, 882–886.

Blume, L., M. Bray and D. Easley (1982), “Introduction to the Stability of Rational Expectations Equilibrium,” *Journal of Economic Theory,* 26, 313–317.

Blume, L. and D. Easley (1984), “Rational Expectations Equilibrium: An Alternative Approach,” *Journal of Economic Theory,* 34, 116–129.

——————————— (1992), “Evolution and Market Behavior,” *Journal of Economic Theory,* 58, 9–40.

——————————— (1993), “What has the Rational Learning Literature Taught Us,” in *Essays in Learning and Rationality in Economics,* ed. by A. Kirman and M. Salmon, Oxford: Blackwell.

Bray, M. and D. Kreps (1987), “Rational Learning and Rational Expectations,” in *Arrow and the Ascent of Modern Economic Theory,* ed. George Feiwel, New York: New York University Press.

Diaconis, P. and D. Freedman (1986), “On the Consistency of bayes Estimates,” *Annals of Statistics,* 14, 1–26.

Doob, J.L. (1949), “Application of the Theory of Martingales,” *Colloq. Internat. CNRS,* 22–28.

Easley, D. and N. Kiefer (1988), “Controlling a Stochastic Process with Unknown Parameters,” *Econometrica,* 56, 1045–1064.

El-Gamal, M. and R. Sundaram (1991), “Bayesian Economists ... Bayesian Agents I: An Alternative Approach to Optimal Learning,” unpublished, California Institute of Technology.

Feldman, M. (1987), “An Example of Convergence to Rational Expectations with Heterogeneous Beliefs,” *International Economic Review,* 28, 635–650.

————————— (1990), “On the Generic Non-Convergence of Bayesian Actions and Beliefs,” *Economic Theory,* forthcoming.

Feldman, M. and A. McLennan (1989), “Learning in a Repeated Statistical Problem with Normal Disturbances,” unpublished, University of Minnesota.

Freedman, D. (1965), “On the Asymptotic Properties of Bayes Estimates in the Discrete Case II,” *Annals of Mathematical Statistics,* 36, 454–456.

Fudenberg, D. and D. Kreps (1988), "Learning, Experimentation and Equilibrium in Games," unpublished, Stanford University.

Fudenberg, D. and D. Levine (1991), "Steady State Learning and Self-Confirming Equilibrium," unpublished, MIT.

Hinderer, K. (1970), *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter,* Berlin: Springer-Verlag.

Jordan, J. (1982), "The Generic Exixtence of Rational Expectations Equilibria in the Higher Dimensional Case," *Journal of Economic Theory,* 26, 224-243.

——— (1991a), "Bayesian Learning in Normal Form Games," *Games and Economic behavior,* 3, 60-81.

——— (1991b), "Bayesian Learning in Repeated Games," unpublished, University of Minnesota.

——— (1992), "Bayesian Learning in Games: A NonBayesian Perspective," unpublished, University of Minnesota.

Jordan, J. and R. Radner (1982), "Rational Expectations in Microeconomic Models: An Overview," *Journal of Economic Theory,* 26, 201-223.

Kalai, E. and E. Lehrer (1990), "Merging Economic Froecasts," unpublished, Northwestern University.

——— (1992a), "Rational Learning Leads to Nash Equilibrium," unpublished, Northwestern University.

——— (1992b), "Subjective Equilibrium in Repeated Games," unpublished, Northwestern University.

Marimon R., E. McGrattan and T. Sargent (1989), "Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents," *Journal of Economic Dynamics and Control,* forthcoming.

McLennan, A. (1987), "Incomplete Learning in a Repeated Statistical Decision Problem," unpublished, University of Minnesota.

Mertens, J.-F. and S. Zamir (1985), "Formalization of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory,* 14, 1-22.

Nyarko Y. (1987), "The Number of Equations Versus the Number of Unknowns: The Convergence of Bayesian Posterior Processes," *Journal of Economic Dynamics and Control,* forthcoming.

——— (1991), "Bayesian Learning Without Common Priors and Convergence to Nash Equilibrium," unpublished, New York University.

——— (1992), "Bayesian Learning in Repeated Games Leads to Correlated Equilibria," unpublished, New York University.

Radner, R. (1972), "Existence of Equilibrium of Plans, Prices and Price Expectations in a Sequence of Markets," *Econometrica,* 40, 289-303.

——— (1979), "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica,* 47, 655-678.

——— (1982), "Equilibrium Under Uncertainty," in *Handbook of Mathematical Economics Vol. 2,* ed. by K.J. Arrow and M.D. Intriligator, Amsterdam: North Holland.

Schwartz, L. (1965), "On Bayes Procedures," *Z. Wahrscheinlichkeitstheorie,* 4, 10–26.

Townsend, R. (1978), "Market Anticipations, Rational Expectations and Bayesian Analysis," *International Economic Review,* 19, 481–494.