On the Structure and Diversity of Rational Beliefs

Mordecai Kurz

SFI WORKING PAPER: 1991-02-013

ON THE STRUCTURE AND DIVERSITY OF RATIONAL BELIEFS

Mordecai Kurz

Stanford University

April 1990

Acknowledgments

This research was supported by NSF Grant IRI-8814954 to Stanford University and by the Santa Fe Institute Economic Research Program which is funded by grants from Citicorp/Citibank and the Russell Sage Foundation, the John D. and Catherine T. MacArthur Foundation, the National Science Foundation (PHY-8714918) and the U.S. Department of Energy (ER-FG05-88ER25054). The author is thankful to Steven N. Durlauf for many stimulating conversations related to this research; to William Brock for extensive comments on earlier draft and to Andrzej Ruszczynski for helpful comments.

On the Structure and Diversity of Rational Beliefs

by

Mordecai Kurz
Stanford University
December 1990
(First Draft: April 1990)

1. Introduction.

The formation of expectations and probability beliefs has played a central role in the formulation of sequential and other dynamic equilibria. The specification of "rational" expectations in either economics or game theory requires agents to possess extraordinary information and knowledge about the underlying structure of the economy or the game. It is usually hard to conceive how agents come to possess such information and knowledge. The recent response to this problem has been to formulate dynamic processes of learning which aim to show how agents learn what they know when formulating their beliefs. The problem is that this research has not solved the initial problem. Without engaging in a full scale survey of the results of the recent effort, we think it is accurate to say that there are examples worked out where complete learning does take place. However, in general, the learning approach has not been able to provide a satisfactory justification for agents to be fully knowledgeable rationally expecting agents. This conclusion has a counterpart in the statistical literaturewhere a spirited debate has been taking place on the problem of "Bayes consistency" (see Diaconis and Freedman [1986] for an excellent recent survey). This is the problem of ensuring the convergence of posterior distributions to the mass-point distribution at the true parameter. We note that "Bayes consistency" may fail even when the statistician is able to conduct independent, repeated controlled experiments. A learning economic agent cannot obtain independent observations and must be content with the acual data generated by the system.

It is interesting to note that both Bayesian as well as rational expectations theories of belief formation have fundamental difficulties in explaining why is it that intelligent economic agents exhibit drastic differences in beliefs without necessarily having substantial differences in the information available to them. Bayesians insist on the common prior assumption (see Aumann [1987] page 12) which requires all agents to have the same prior given that they have the same information. We take the existence of diversity of beliefs among intelligent and equally informed agent to be an empirical observation which requires an explanation.

The prototype problem with which this paper is concerned may be simply explained with the aid of an example. Let y_t, $t = 0,1,2,\ldots$ be a sequence of random profits or rewards of a household, a corporation or an investment project. Let $0 < \gamma < 1$ be the discount rate employed and let the present value, at date t, of future rewards be defined by

(1) $$p_t^* = \sum_{k=0}^{\infty} \gamma^{k+1} y_{t+k} .$$

An economic agent who observes the data needs to evaluate, at date t, the risky prospect p_t^* via its distribution or moments like Ep_t^*, Var p_t^* etc. The problem is that the agent does not know the true probability distribution of the random sequence $\{y_t, t = 0,1,2,\ldots\}$. He does have a massive amount of past data since $t = 0$ occurred a long time ago and all past data was recorded. Given this data the agent sets up to learn all that he can and then form a conditional probability belief Q^t about the future sequence of random variables y_{t+k}, $k = 0,1,2,\ldots.$ from date t on. With this probability selected, the agent can compute the implied distribution of p_t^* and its moments. We aim to establish criteria to determine if a probability belief of an agent is "rational." Moreover, given such criteria of rationality, can we explain why two rational agents endowed with the same information may come up with drastically different beliefs?

The problem of evaluating future risky events is made particularly difficult since no market for contingent claims exist. This is a consequence of many well known reasons: the "state" is not observable, many imputs such as labor cannot be legally traded on futures markets, futures contracts cannot be enforced, many crucial components of the "state" suffer from problems of moral hazard and other incentive effects, etc. With incomplete market structure arbitrage considerations are of limited significance. We are explicitly rejecting the Rational Expectations theory and the reasons will become evident from the development below.

Returning now to our discussion on learning procedures it is important to note that one central achievement of this literature is that it established the basic requirement that a theory of the formation of beliefs must be based on the knowledge which agents actually have rather than on a hypothetical knowledge which the model builder may want them to have. This calls, however, for the precise identification of what is knowable by the agents and what is not knowable. Since the characterization of the unknowable is known to all agents, such recognition must have important implications.

This paper explores the process of the formation of beliefs by starting from the suggested view that such beliefs must be compatible with what is learnable from the data. We formulate the economic environment as a stochastic dynamical system in which all agents know the structure of the system but they do not know the probability which governs the stochastic process. This true probability is the central object of learning. Agents can observe all past data generated by the system and we postulate that the amount of such past data is very large. It will become immediately clear that there are such erratic economies in which nothing can be learned and others, in which everything is learnable. We shall confine our attention to those economies that exhibit stability properties which enable some learning. Stability is defined in terms of the convergence of the long term relative frequencies at which events occur. The assumption of stability, but not necessarily stationarity, will enable us to argue that all the agents will be able to learn a stationary (i.e. invariant) probability of the dynamical system. We shall insist that this is all that the agents can ever learn and therefore this should be taken as the primitive empirical knowledge.

In addition to the empirical knowledge which all agents share, our agents face imperfect knowledge about the structure of the economy and the nature of the random mechanism which generates the data. This imperfection leads to the formation of conjectures, hypotheses, theories and other forms in which intelligent opinions may be expressed. We hold the view that randomness in the economy is a real phenomenon determined by the structure of the economy, its organization and the technology at its foundation. A better scientific understanding of this structure will lead to an improved knowledge of the causes of random fluctuations. Obviously, all theories must be tested with the aid of the data generated by the system. The problem is that often the data provides only a partial resolution of the differences among competing theories. This leaves a wide scientific gap which can be resolved only by further scientific developments. It is these scientific gaps which give rise, at any moment in history, to wide diversity of opinions. In this paper we study the formation of beliefs <u>given</u> a state of human knowledge.

We take the data generated as the basis for a definition of "rationality of belief". We postulate two axioms which all rational beliefs must satisfy. The main theorem of the paper provides a characterization of all beliefs which satisfy the axioms. The characterization shows that a rational probability belief must be a convex combination of two probabilities: one is equivalent to the stationary, invariant, measure and the other is orthogonal to it. This characterization helps clarify the observation that the diversity of beliefs among intelligent agents with the same information is the consequence of their common understanding that gaps exist in their knowledge which cannot be resolved with the available data.

The paper provides a series of propositions and examples which aim to clarify the basic ideas.

2. The Model and Basic Formulation.

An abstract representation of our economic environments specifies the economy as a dynamical system (Ω, F, Π, T). Here Ω is a subset of a complete and separable metric space, F is the σ-field of the Borel subsets of Ω, Π is a probability measure on the measurable space (Ω, F) and $T: \Omega \to \Omega$ is a measurable transformation. In more concrete terms we postulate that an economic agent observes a vector $x_t = (x_{1t}, x_{2t}, \ldots, x_{kt}) \in X$ of random variables with X being the state space. Normally the x_{it} will be such quantities as the profits of firms, outputs of firms or industries, prices of assets and commodities, climate conditions etc. Hence, for all practical purposes we may as well assume $X \subset \Re^k$. We denote by x a generic infinite sequence in $(\Re^k)^\infty$. We shall assume that economic agents know the postulated structure but do not know the distribution of x in $(\Re^k)^\infty$. We can attain a simplification of notation if we use the fact that all countably generated probability spaces are isomorphic (see Parthasarathy [1967] Chapter I). This allows us to define (Ω, F, Π) as the coordinate probability space:

$$\Omega = X^\infty \subset (\Re^k)^\infty$$

$$F = B(X^\infty) = \text{the } \sigma\text{-field of Borel sets of } X^\infty$$

and Π is a probability on measurable sets of infinite sequences in X^∞. Although we shall think of x as a random point in X^∞ it will be important for us to associate random points x with the time at which they are selected. We use the notation $x^t = (x_t, x_{t+1}, \ldots)$ to identify a random sequence at time t. This brings up one possible confusion: sometimes we would want to talk about $x^t = (x_t, x_{t+1}, \ldots)$ as the sequence of random variables observable from time t. Similarly we may want to talk about x as an infinite sequence of random variables in X^∞. The context will usually allow this distinction but for this reason we shall preserve the notation of both x and w keeping in mind the identification $\Omega = X^\infty$. Whenever we use the notation ω we shall be stressing the fact that we are looking at a particular sample point $\omega \in X^\infty$.

The realization of the stochastic process is given by a dynamical structure which is represented by the transformation T. If the realization at time t is $x^t \in X^\infty$ and at time $t+1$ it is $y^{t+1} \in X^\infty$ then

$$y^{t+1} = Tx^t . \tag{1}$$

It is very common to think of the random sequence $x = (x_0, x_1, x_2, \ldots) \in X^\infty$ as being selected simultaneously. In such a case $x^t = (x_t, x_{t+1}, x_{t+2}, \ldots)$ is the sequence of realizations at dates t. This gives rise to a definition of the shift transformation:

Definition 1: A measurable transformation T is said to be a shift transformation if for all $x^t = (x_t, x_{t+1}, \ldots) \in X^\infty$

$$Tx^t = x^{t+1} = (x_{t+1}, x_{t+2}, \ldots) \in X^\infty. \tag{2}$$

Although we do not need to assume that T is a shift transformation such an assumption can be made with a minimal loss of generality. In almost any example or an application we shall think of T as such a transformation.

One defines $T^2x = T(Tx)$ and, in general, $T^n x = T(T^{n-1}x)$. From the measurability of T it follows that the iterated maps T^n are also measurable transformations. We shall assume that the process starts at a date called $t = 0$ with $x^0 = x$. We therefore define

(3) $\qquad y^t = T^t x \qquad t = 0,1,2,\ldots$

The reader may note that we permit t to take values in the set of non-negative integers $\{0,1,2,\ldots\}$. We specifically <u>do not permit negative integers</u>. The force of this assumption is that the transformation T is <u>not assumed to be invertible</u>. The economic meaning of this assumption is that any particular <u>future</u> evolution, x^t, of the economy is not associated with a <u>unique past</u> $T^{-1}(x^t)$; a future x^t may arise from many possible pasts! We shall return to this question later.

The above discussion leads to a very important notational convention employed in the paper. Since T is not assumed to be invertible we reserve the notation $T^{-n}S$ to be the preimage of S under T^n. That is

$$T^{-n}S = \{x: \; T^n x \in S\}.$$

For this reason we think of a set $B = \{x \in X^\infty, \; T^n x \in S\}$ as the set $S \subset X^\infty$ located n periods into the future. Since $T^n x \in S$ for all $x \in B$, it follows that B is the set of points in X^∞ from which one reaches S in n steps.

Our objective is to study what an observer can conceivably learn from the data generated by the process. As stated earlier the observer knows the basic structure $(\Omega, \mathcal{F}, \Pi, T)$ except for the probability Π. He can learn something about Π only to the extent that the dynamical system exhibits empirical repetition and, in addition, sufficient data is available to discover any repetitive regularity. Since the objective of the observer is to discover, for each measurable set $S \in \mathcal{F}$, the probability $\Pi(S)$ he has a natural way to proceed. He can define

$$1_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

and then compute

$$m^n(S)(x) = \frac{1}{n} \sum_{k=0}^{n-1} 1_S(T^k x). \tag{4}$$

$m^n(S)(x)$ is the relative frequency at which the dynamical system visits the set S given that it started at x. Our observing agent can conceivably learn something about the true Π only if $m^n(S)(x)$ converges so that with sufficient data $\lim_{n\to\infty} m^n(S)(x) = m(S)(x)$ can be computed to any desired accuracy. This motivates

<u>Definition 2</u>: A dynamical system (Ω, F, Π, T) is said to be <u>stable</u> if for all $S \in F$ the limit of $m^n(S)(x)$ exists for Π almost all x and we then write

$$\lim_{n\to\infty} m^n(S)(x) = m(S)(x) \quad \Pi \text{ a.e.} \tag{5}$$

Note that since $0 \le m^n(S)(x) \le 1$ the lack of convergence of $m^n(S)(x)$ means that for increasing lengths of time the means $m^n(S)(x)$ remain in different parts of the interval $[0,1]$ without ever settling down. It is hard to visualize what one can learn about $\Pi(S)$ from observing such a sequence. We thus propose that in order to be able to talk about "learning" the probability of an event S it must be true that the limit of the empirical frequency of that event exists and agents have enough data to calculate it approximately. In the development below we shall assume that the limits in (5) are known to all agents and this creates somewhat of a dilemma; on the one hand we insist on starting the observations at some date $t = 0$. On the other hand we assume that the agent has as much information as he needs to calculate $m(S)(x)$ for all $S \epsilon F$. Since the assumption that the limits $m(S)(x)$ are known to all the agents is central to our theory, it requires some discussion.

It is important to keep in mind, at the outset, that we are concerned here with on-going economic activity for which substantial history is available. In terms of our model this means that the starting date $t = 0$ for that activity is far in the past. With a substantial amount of past data available, it is possible to obtain an approximation of the limits $m(S)(x)$ to a high degree of accuracy. The idea of approximating $m(S)(x)$ by large, finite, set of data is strongly supplemented by our second observation that in all economic applications agents discount the future and are, therefore, concerned only with events which will occur within a finite horizon. For example, consider the present value of a stream of future profits introduced in (1) above

$$p_t^*(y) = \sum_{k=0}^{\infty} \gamma^{k+1} y_{t+k} .$$

In practical applications it is difficult to think of problems where profits 20 years after date t will make a significant difference to $p_t^*(y)$. Keeping this in mind, let the σ-field of J horizon events at date t be defined by

$$F^{t,t+J} = \sigma(x_t, x_{t+1}, \ldots, x_{t+J})$$

and let

$$F^{t,\infty} \equiv F^{t} = \sigma(x_t, x_{t+1}, \ldots) \quad .$$

Then, for any problem there exists a finite J such that economic agents would be concerned mostly with the probability of J horizon events. In any practical situation, the error in economic values of using sets in $F^{t,t+J}$ instead of F^t, is negligible.

Consider now any J horizon set S in $F^{0,J} = \sigma(x_0, x_1, \ldots, x_J)$. If n is large relative to J and if $m^k(S)(x)$ converges sufficiently rapidly then $m^n(S)(x)$ will provide a good approximation to its limit $m(S)(x)$. Under these circumatances our assumption that $m(S)(x)$ is known amounts to the assumption that the history of the process is sufficiently long to enable the agents to learn the normal dynamical patterns of the process over time intervals of length J which are relevant to their welfare.

Our third point is that $m(S)(x)$ may be deduced from the underlying economic model which generates the data. In general, economic theory provides a better explanation of long-term tendencies than of short-term fluctuations. For example, considerations of equilibrium and free entry combined with long-term resource and technological patterns, may often provide a reasonable basis for a judgment of what average long-term patterns of output, profits, etc. should be. In our analysis $m(S)(x)$ describe these long-term patterns and therefore may be deduced from general economic theoretic considerations.

Our fourth point is a methodological one. As we pointed out above, one should think of the limits $m(S)(x)$ as the average or normal patterns of the dynamical system. Our aim is to study the formation of beliefs based on what is conceivably knowable by the agents. In practice, one develops algorithms to approximate $m(S)(x)$ with large finite data but this does not alter the basic fact that it is $m(S)(x)$ which all agents can conceivably learn. The idea of endowing the agents with what they can conceivably learn is a methodological simplification which we are making in order to avoid the complication of approximations.

When one considers economic variables, there are general technological and physical limitations imposed on the values which they may take: the output of wheat cannot exceed what the entire world may produce and the output of steel cannot exceed twice the rated capacity. We <u>do not</u> propose to assume that for any $x \in X^\infty$ we have that x_t is known by all agents to be uniformly bounded. Instead we shall assume that Π is tight and all agents know that it is tight. Given a probability space $(\Omega, \mathcal{F})$ where Ω is a complete and separable metric space a probability measure Π is said to be <u>tight</u> if for each $\epsilon > 0$ there exists a compact set $K_\epsilon \subset \Omega$ such that

$$\Pi(K_\epsilon) > 1 - \epsilon .$$

Thus, on general scientific principles agents know that Π is tight but they may have their own opinions of what K_ϵ is for each ϵ.

<u>Summary of Assumptions.</u>

The economy is described by a dynamical system $(\Omega, \mathcal{F}, \Pi, T)$ defined on $t = 0,1,2,\dots$.

<u>Assumption 1</u>: $\Omega = X^\infty$ where $X \subset \Re^k$ and

$\mathcal{F} = B(X^\infty) =$ the σ-field of Borel subsets of X^∞.

<u>Assumption 2</u>: The dynamical system $(\Omega, \mathcal{F}, \Pi, T)$ is stable and agents know $m(S)(x)$ for all $S \in \mathcal{F}$ and for all $x \in \Omega$.

Assumption 3: The dynamic process takes place on the set $\{0,1,2,\ldots\}$ of non-negative integers. T is not necessarily invertible.

Assumption 4: Agents know the description $(\Omega,\mathcal{F},\Pi,T)$ but they do not know Π. The probability Π is tight and all agents know this. However, the tightness conditions are agent specific (i.e. for each $\epsilon > 0$, K_ϵ is agent specific).

3. Stability and Asymptotically Mean Stationarity.

3.1 Stationary Systems.

An important case where a dynamical system has adequate repetition is the case of a stationary system. The dynamical system $(\Omega,\mathcal{F},\Pi,T)$ is said to be stationary if the transformation T is measure preserving; that is, if for all $S \in \mathcal{F}$

$$\Pi(T^{-1}S) = \Pi(S).$$

When T preserves Π then Π is said to be invariant under T.

Almost all results in Ergodic Theory have been proved for the case of measure preserving transformations. When a dynamical system is stationary and agents know that it is stationary the questions raised in this paper have very clear answers. The main tool employed is Birkhoff's Ergodic Theorem (also known as the "Pointwise" Ergodic Theorem or "The" Ergodic Theorem). Since we use various aspects of this theorem it will be useful to state it and see later some of its implications. We note first that in calculating the relative frequency function $m^n(S)(x)$ in Equation (4) the agents use the characteristic function $1_S(x)$. Since the ergodic theorem holds with respect to a broader class of functions we define, for any measurable function, the average

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x).$$

Now we introduce the following terms:

Definition 3: A measurable set S is said to be invariant with respect to T if $T^{-1}S = S$. A measurable function is said to be invariant with respect to T if for any $x \in \Omega$, $f(Tx) = f(x)$.

Definition 4: A dynamical system is said to be ergodic if $\Pi(S) = 0$ or $\Pi(S) = 1$ for all invariant sets S.

Now let the collection I of invariant sets be defined by

$$I = \{S \in F \ : \ T^{-1}S = S\}.$$

It is easily seen that I is a sub σ-field of F and hence one can define the conditional probability of Π given I; we denote it by

$$\Pi(S|I)(\omega) \quad \text{for all} \quad S \in F,\ \omega \in \Omega.$$

The Ergodic Theorem (Birknoff (1931): Let (Ω, F, Π, T) be a stationary dynamical system and let the measurable function $f \in L^1(\Omega, F, \Pi)$. Then

(i) $$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \bar{f}(x) \quad \text{exists a.e.},$$

(ii) $\bar{f} \in L^1(\Omega, F, \Pi)$ is an invariant function,

(iii) $$\overline{f}(x) = E_{\Pi}(f|I)(x) \quad \text{a.e.},$$

(iv) if the dynamical system is ergodic then

$$\overline{f}(x) = \overline{f} = E_{\Pi}f \quad \text{a.e.}$$

Applying the ergodic theorem to our problem when (Ω,F,Π,T) is stationary, we can draw three direct implications:

(a) $\lim_{n\to\infty} m^n(S)(x) = m(S)(x)$ exists for all $S \in F$ a.e.,

(b) $m(S)(x) = \Pi(S|I)(x)$ for all $S \in F$ a.e.,

(c) if (Ω,F,Π,T) is ergodic then

$m(S)(x) = m(S) = \Pi(S)$ for all $S \in F$ a.e.

It is then clear that if the dynamical system is stationary <u>and the agents know that it is stationary</u> then they can calculate $m(\cdot)(x)$ and know that they have learned exactly the conditional probability $\Pi(\cdot|I)(x)$. In the ergodic case the agents calculate the measure m and know that $m = \Pi$. The conclusion that $m(S)(x) = \Pi(S|I)(x)$, $S \in F$, in the non-ergodic case is sensible since in this case the sequence $(T^n x)$ will visit only the invariant sets which contain x and hence $\Pi(\cdot|I)(x)$ is the only object which can be learned.

It is useful to point out that when the dynamical system is stationary agents may not know that it is stationary. Moreover, there does not exist any statistical means by which agents can ascertain that a stationary system is, in fact, stationary. More important is the fact that the dynamical system may not be stationary. In this eventuality, even if we work with a stable system for which $m(\cdot)(x)$ exists, agents cannot use the ergodic theorem--as stated--to be able to determine what is it that they are learning.

In our view the determination if a dynamical system is stationary or not must originate with the foundation of the model which gives rise to the system. Two examples will illustrate the point. In certain applications in physics the description of stochastic dynamical systems arise from Hamiltonian structures. These Hamiltonians imply that the transformation of the dynamical systems are measure preserving and thus stochastically stationary. However, this stationarity of the transformation can be traced to the fact that Hamiltonians are required to satisfy Liouville's theorem on the conservation of energy. Putting it differently, the stationarity of the dynamical system is proved as a logical consequence of Liouville's Theorem which, in turn, is proved from the underlying physical structure. Thus, stationarity is a logical implication of the underlying theory rather than an empirical observation which is deduced from the data.

In statistical applications, consider the sequence of observations x_t for $t = 0,1,2,\ldots$ where $x_t = 1$ or $x_t = 0$ depending upon the result of an experiment, like the toss of a coin, in which two outcomes are possible with probability p and $q = 1 - p$. If the experiments are conducted independently, then it is a logical implication that the stochastic process $\{x_t, t = 0,1,2,\ldots\}$ is stationary and, considering the implied measure on sets of infinite sequences $x \in \Re^{\infty}$, the shift transformation preserves the measure. Here again, the claim that a system is stationary originates from the logical structure of the stochastic mechanism which generates the data.

It is, perhaps, the fundamaental starting point of this paper that in economic systems we rarely have more than a superficial knowledge of the stochastic factors which generate the data. Equilibrium considerations often help us understand the "typical" or "normal" fluctuations of a random economic system but very little in economic theory is designed to enable us to make a logical deduction of what must be the nature of the probability laws under which the economic data is generated. More specifically, we suggest that although the assumption of stationarity is almost universally employed in applied economics, there is little theoretical justification for it. If anything, there are compelling theoretical reasons to expect economic systems to be non-stationary.

3.2 Why Non-Stationarity.

A discussion of the evidence for non-stationarity in dynamic economic processes takes us into a very broad arena. Our aim here is not to provide a comprehensive survey of the evidence but rather, to state briefly the arguments against an a-priori presumption of stationarity.

(i) Technological Changes. Economic growth has been associated with dramatic changes in technology, product mix and human knowledge. Most empirical evidence suggests that the pattern of these changes is complex and irregular.

(ii) Externalities, Returns to Scale and Non-Convexities. Most students of economic growth agree that non-convexities play an important role in the dynamics of growth. An extremely large literature which has evolved over the last two-three decades argues that these non-convexities result in path-dependency, critical sensitivity to changes in basic parameters and multiple equilibria.

(iii) The Persistent Impact of Major Events. Almost every empirical work with time series suggests that the data contains short periods of sporadic behavior which leave their marks for a long time. Depressions, wars, crashes, speculative bubbles and other such phenomena have lasting effects and remain outliers no matter what stationary model is employed.

(iv) Unobservability of the State and Aggregation. Suppose that the "state" of an economic system is described by a vector $y = (y_1, \ldots, y_M) \in \Re^M$. Given that technology, production, resource availability, preference of agents, levels of effort and other measures of incentives are all random variables included in y, it is obvious that we are dealing with an extraordinarily large dimension M of, essentially, unobservable variables. Instead we observe a much smaller number of either aggregates or other transformed variables such as GNP, output of industry j, profits of firm k etc. What must be strongly noted is that even if $\{y_t = (y_{t1}, y_{t2}, \ldots, y_{tM});\ t = 0,1,2,\ldots.\}$ is a stationary process, this stationarity is not preserved under aggregation or under other measurable transformations! Since non-observability of the state is a major cause for incompleteness of the market structure, it appears that there may exist an intimate connection between the non-stationarity properties and the incomplete nature of a dynamical market economy.

The combination of all the factors reviewed in (i) - (iv) above suggests to us that in the absence of conclusive theoretical reasoning which proves that a process must be stationary, we would expect that rational beliefs will not reject the possibility that the process is non-stationary.

3.3. Stable Systems and the Ergodic Theorem.

Returning now to our main theme, we insist that our dynamical system (Ω, F, Π, T) may not be stationary. However, in view of the ergodic theorem one must immediately ask if the conditions of stability and non-stationarity are compatible? To examine this question we return to the definition of stability.

The condition of stability requires that for any measurable set $S \in F$ we must have

$$\lim_{n\to\infty} m^n(S)(x) = m(S)(x) \quad \text{exists} \quad \text{a.e.}$$

Since $0 \leq m^n(S)(x) \leq 1$ and $m^n(S)(\cdot)$ is a finite sum of measurable functions on Ω, $m^n(S)(\cdot)$ is also a bounded measurable function. Taking expectations on both sides and passing to the limit yields that

$$\lim_{n\to\infty} \int_\Omega m^n(S)(x)\Pi(dx) = \int_\Omega m(S)(x)\Pi(dx) \quad . \tag{6}$$

However

$$\int_\Omega m^n(S)(x)\Pi(dt) = \frac{1}{n}\sum_{k=0}^{n-1} \int_\Omega 1_S(T^k x)\Pi(dx).$$

Since

$$1_S(T^k x) = \begin{cases} 1 & \text{if} \quad x \in T^{-k}S \\ 0 & \text{if} \quad x \notin T^{-k}S \end{cases}$$

it follows that

$$\int_\Omega 1_S(T^k x)\Pi(dx) = \Pi(T^{-k}S).$$

This proves that the expectations of $m^n(S)(\cdot)$ is the mean probability, which means

$$(7) \qquad \int_{\Omega} m^n(S)(x)\Pi(dx) = \frac{1}{n}\sum_{k=o}^{n-1}\Pi(T^{-k}S).$$

Combining (6) and (7) we can conclude that the stability of a dynamical system requires that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{n=0}^{n-1}\Pi(T^{-k}S) = m(S) \quad \text{exists for all} \quad S \in F .$$

This motivates our next definition

<u>Definition 5</u>: A dynamical system (Ω, F, Π, T) is said to be <u>asymptotically mean stationary</u> if for all $S \in F$ the limit

$$(8) \qquad m(S) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\Pi(T^{-k}S) \quad \text{exists.}$$

Although asymptotically mean stationary processes are, generally, not stationary, it turns out that many of the tools of ergodic theory remain applicable to them. Such processes were studied by Dowker [1951], [1955], Rechard [1956], Gray and Kieffer [1980] and Gray [1988]. Also, some studies in Information Theory have employed such processes (see for example Fontana, Gray and Kieffer [1981] and Kieffer and Rahe [1981]). We shall make extensive use of these results.

The first observation to be made is that if we denote the mean probability of S by

$$\Pi^n(S) = \frac{1}{n} \sum_{k=0}^{n-1} \Pi(T^{-k}S) \tag{9}$$

then Π^n is also a probability on (Ω,F). The definition of asymptotically mean stationary processes requires the sequence Π^n to converge to some m in the sense that for each $S \in F$ $\Pi^n(S) \to m(S)$. It then follows from the Vitali-Hahn-Saks theorem (see Neveu [1965], page 117) that m is also a probability on (Ω,F). Moreover, the measure m is invariant relative to T and hence the dynamical system (Ω,F,m,T) is stationary. To see why T preserves m we calculate

$$\begin{aligned}
\Pi^n(T^{-1}S) &= \frac{1}{n} \sum_{k=0}^{n-1} \Pi(T^{-k}(T^{-1}S)) \\
&= \frac{1}{n} \sum_{k=0}^{n} \Pi(T^{-k}S) - \frac{1}{n}\Pi(S) \\
&= \frac{1}{n}\Pi(S) + \left(\frac{n+1}{n}\right) \frac{1}{n+1} \sum_{k=0}^{n} \Pi(T^{-k}S) \\
&= -\frac{1}{n}\Pi(S) + \frac{n+1}{n}\Pi^{(n+1)}(S).
\end{aligned}$$

Taking limits as $n \to \infty$ we can then conclude that

$$\lim_{n\to\infty} \Pi^n(T^{-1}S) = m(T^{-1}S) = \lim_{n\to\infty} \Pi^{(n+1)}(S) = m(S)$$

hence $m(T^{-1}S) = m(S)$. m is a stationary or invariant probability measure of (Ω,F,Π,T).

The second observation to be made is central to our development:

Theorem 1: (Dowker [1951], Rechard [1956]): A dynamical system (Ω, F, Π, T) is stable if and only if it is asymptotically mean stationary.

Theorem 1 shows that the condition of asymptotically mean stationarity is exactly the natural condition for economic environments in which some learning can take place but without necessarily being stationary. The central result of the theory of asymptotically mean stationary processes says that the condition of stability of a dynamical system is sufficient for the validity of the ergodic theorem:

Theorem 2: Let (Ω, F, Π, T) be stable with a stationary measure m. If f is a measurable function with $f \in L^1(\Omega, F, m)$ then

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^{-k}x) = \bar{f}(x)$ exists Π a.e.

(ii) $\bar{f} \in L^1(\Omega, F, m)$ is an invariant function.

(iii) $\bar{f}(x) = E_m(f|I)(x)$ Π a.e.

(iv) if (Ω, F, Π, T) is ergodic then

$$\bar{f}(x) = \bar{f} = E_m f \quad \text{independent of } x \qquad \Pi \text{ a.e.}$$

(v) in the special case of $f = 1_S$, $S \in F$, we have

$$\lim_{n\to\infty} m^n(S)(x) = m(S)(x) = m(S|I)(x) \quad .$$

If the dynamical system is ergodic then $m(S)(x)$ and $m(S|I)(x)$ are independent of x and therefore $m(S|I)(x) = m(S)$ Π a.e.

Theorem 2 shows that when a dynamical system (Ω, F, Π, T) is stable with a stationary measure m then it generates, for almost each $x \in X^\infty$ and

$S \epsilon F$, a limiting empirical relative frequency $m(S)(x)$ which is equal to $m(S|I)(x)$ - the m conditional probability of S given the σ-field of T invariant events. If the dynamical system is ergodic then $m(S)(x)$ is independent of x and equals $m(S)$. Theorem 2 also helps clarify the nature of our earlier Assumption 2. This assumption states that the limits in (5) and known to the agents. Theorem 2 shows that this is equivalent to the knowledge of the invariant measure m. For this reason we shall restate Assumption 2 in the following manner:

Assumption 2': The dynamical system (Ω, F, Π, T) is stable and all agents know the associated invariant measure m.

4. Trends and Periodicity.

Restricting attention only to stable and tight dynamical systems may appear to be excessively narrow in view of the existence of trends and periodicity in economic time series. Starting with trends we note first that such economic variables as output, employment, profits of firm j etc., all have trends and could thus pass the test of stability only in a somewhat unsatisfactory way: for any bounded set S we shall have $m(S) = 0$ which is in violation of Assumption 4 on tightness. Recall that our aim is to establish criteria for rationality of beliefs. In economic applications this ultimately requires us to specify, at each date t, the conditional probability of a future event S given the past $(x_0, x_1, \ldots, x_t)$. For this it is sufficient to specify the probability of increments from x_t into any set S of future values of the variables. The typical procedure of handling this is to transform the variables. If, on the average, the growth rate is geometric then $y_t = x_t/x_{t-1}$ will do.

If the growth is linear then $y_t = x_t - x_{t-1}$ will do. For other growth patterns appropriate measurable transformations may be considered. In general it is the case that if $\{x_t, t = 0,1,2,...\}$ is stable, a measurable transformation would imply that $\{y_t, t = 1,2,...\}$ is also stable with a stationary measure m_y. The problem is that m_y may not be tight. We note in passing that the assumption of stability is reasonably mild while the tightness condition imposes substantial regularity on the measure.

If a transformation can be found so that m_y is tight then all the assumptions of this paper are satisfied with respect to the process $\{y_t, t = 1,2,...\}$. The agent can now derive the implied conditional probability of future x events given $(x_0, x_1, ..., x_t)$.

Turning now to periodicity let us start the discussion with an example that could help clarify the main ideas.

<u>Example 1</u>. consider a process $\{x_1; t = 0,1,2...\}$ of independent random variables $x_t \in \mathbb{R}^1$ specified as follows:

x_t is distributed uniformly on $[0,1]$ if t is even.

x_t is distributed uniformly on $[\frac{1}{2}, 1\frac{1}{2}]$ if t is odd.

This is clearly not a stationary process. Consider now the following measurable set:

$$S_1 = \{y \in X^\infty: \quad 0 \leq y_1 \leq \frac{1}{2}\}$$

where y_1 is the first coordinate of y (thus if $y = (x_t, x_{t+1}, x_{t+2}, ...)$ then $y_1 = x_t$). Given the observations $T^t x$ the agent will calculate

$$\frac{1}{n} \sum_{k=0}^{n-1} 1_{S_1}(T^k x).$$

This procedure will average between 50% of the time in which the draw is uniform on $[\frac{1}{2}, 1\frac{1}{2}]$ and 50% of the time in which the draw is uniform on $[0,1]$. It then follows that

$$m(S_1)(x) = \frac{1}{4} \qquad \text{a.e.}$$

Next the agent may consider the measurable set

$$S_2 = \{y \in X^\infty: 0 \le y_1 < \frac{1}{2}, 0 \le y_2 \le \frac{1}{2}\}.$$

It is easily seen that

$$m(S_2)(x) = 0 \qquad \text{a.e.}$$

A further examination will reveal that for a set like

$$S_3 = \{y \in X^\infty: \frac{1}{2} \le y_1 \le 1\frac{1}{2}, 0 \le y_2 \le 1\}$$

the calculation yields

$$m(S_3)(x) = \frac{5}{8}$$

and for a set like

$$S_4 = \{y \in X^\infty: 0 \le y_1 \le 1, \frac{1}{2} \le y_2 \le 1\frac{1}{2}\}$$

the calculation yields

$$m(S_4)(x) = 5/8 \qquad \text{a.e.}$$

These facts and others will soon convince the learning agent that in search of stationarity he must think of the data as a sequence

$$x^{2t} = (x_{2t}, x_{2t+1}, x_{2t+2}, \ldots) \quad t = 0,1,2,\ldots$$

In this process the dynamical system transforms x^t in one period into x^{t+2} rather than into x^{t+1}. Equivalently, it transforms $x = (x_0, x_1, x_2, \ldots)$ into x^{2t} in t periods rather than into x^t. One way of handling this is by considering the dynamical system to be (Ω, F, Π, T^2) and hence if a vector $x \in X^\infty$ is given then

$$T^2 x = (x_2, x_3, x_4, \ldots)$$

$$\vdots$$

$$T^{2t} x = (x_{2t}, x_{2t+1} \cdot x_{2t+2}, \ldots).$$

In the case of the example above the agent can calculate the stationary probability $m_2(\cdot)(x)$ which is derived from T^2. For the sets (S_1, S_2, S_3, S_4) this yields

$$m_2(S_1)(x) = 1/2$$

$$m_2(S_2)(x) = 0$$

$$m_2(S_3)(x) = 1/4$$

$$m_2(S_4)(x) = 1 \ .$$

Examination of T^j for $j = 1,2,3,\ldots$ reveals that the stationary measures $m_j(\cdot)(x)$ satisfy, for all $S \in F$

$$m(S)(x) \equiv m_1(S)(x) = m_3(S)(x) = m_5(S)(x) = \ldots$$

$$m_2(S)(x) = m_4(S)(x) = m_6(S)(x) = \ldots$$

and also that

$$m(S)(x) = \frac{1}{2}\, m_2(S)(x) + \frac{1}{2}\, m_2(T^{-1}S)(x) \quad .$$

This will convince the agent that the system has periodicity of 2 and no more.

The results of the example can be generalized. Consider any dynamical system (Ω,F,Π,T) and its associated "N block" system (Ω,F,Π,T^N). It can be shown that if (Ω,F,Π,T) is stable with stationary measure m then (Ω,F,Π,T^N) is also stable with a stationary measure m_N. Moreover, the relationship between m and m_N satisfies for all $S \in F$

$$m(S) = \frac{1}{N} \sum_{j=0}^{N-1} m_N(T^{-j}S) \quad .$$

Note that it is T^N which preserves m_N and not T consequently $m_N(T^{-j}S) = m_N(S)$ only if $j = N$. It is then clear that if the dynamical system has no periodicity then $m_k = m$ for all k. On the other hand we say that the system has periodicity if for some k, $m_k \neq m$.

It is essential to see that the discovery that a system may have periodicity says nothing about its stationarity. In the example above the system is stationary when we consider it to be a random sequence of pairs $z_t = (x_{2t}, x_{2t+1})$ $t = 0,1,2,\dots$. However, the examination of the blocks reveals the periodicity through the fact that for all T^j there exist only two distinct stationary measures:

$$m = m_j \qquad \text{for} \quad j \quad \text{odd}$$
$$m_2 = m_j \qquad \text{for} \quad j \quad \text{even.}$$

This discussion leads us to conclude that the problems of trend and periodicity are only technical in nature. Having clarified their nature we shall assume in the balance of this paper that these features are absent from the process under study.

5. The Structure of Rational Beliefs.

Our aim now is to provide a characterization of rational beliefs. Since in this paper all beliefs are formed only after observing a great deal of data, the concept of "rationality" must be understood with respect to statements about conditional probabilities of future events given the observed past data. The key question is what is the relevant empirical knowledge which should dictate to a rational agent his choice of conditional probabilities of future events. In the previous sections we have endeavored to show that the relevant empirical knowledge is entirely represented by the stationary probability measure m and all rational beliefs should be required to be compatible with this knowledge.

The approach adopted in this paper is based on Ergodic theory rather than on Bayesian statistics and for this reason it is not convenient to specify criteria for the selection of rational conditional probabilities. Instead we specify two Axioms of Rationality under which the agent selects a probability P which he believes to be the true measure on $(\Omega,\mathcal{F})$. Since our central interest is in conditional probabilities the reasonableness of our procedures and Axioms should be evaluated relative to the implied selection of the conditional probability $P^t(S|x_{(t)})$, $S \in \mathcal{F}$. Our justification is simple: at the given fixed date t at which the agent forms his belief he is assumed to know both the stationary measure m of $(\Omega,\mathcal{F},\Pi,T)$ as well as the actual past data $x_{(t)} = (x_0,x_1,\ldots,x_{t-1})$

generated by the dynamical system; all economic choices at date t will depend upon the conditional probability $p^t(\cdot|x_{(t)})$ only.

5.1. Forming Beliefs

We presume that each agent makes explicit or implicit probability assessment of events in Ω. Thus, simply stated, the formation of a belief by an agent about the dynamical system $(\Omega,\mathcal{F},\Pi,T)$ must result in his selecting a probability P and then presuming that the dynamical system is $(\Omega,\mathcal{F},P,T)$. This means that when the agent faces a decision problem over time he will assign the probability $P(S)$ to each event $S \in \mathcal{F}$. P is the probability which the agent will use to evaluate uncertain prospects.

To implement the above viewpoint recall that our agent is given a stationary measure m on $(\Omega,\mathcal{F})$ which was generated by $(\Omega,\mathcal{F},\Pi,T)$; he is then asked to form a belief about the true Π. From this vantage point if $\mathcal{P}(\Omega)$ is the set of all probabilities on Ω, then the object of uncertainty is $\mathcal{P}(\Omega)$ itself and hence <u>forming a belief</u> about Π would necessitate <u>the agent's selecting a probability P^* on $\mathcal{P}(\Omega)$</u>. That is, if $\mathcal{P}(\mathcal{P}(\Omega))$ is the space of all probabilities on $\mathcal{P}(\Omega)$ then the agent must select $P^* \in \mathcal{P}(\mathcal{P}(\Omega))$. P^* is an agent specific or "subjective" probability over the set of all possible probabilities which the agent may adopt for his decision-making. However, given the need to select such a probability we propose to make this selection by simply defining

$$P = \int_{\mathcal{P}(\Omega)} \mu P^*(d\mu) \tag{10}$$

if such an integral makes sense. Under this procedure we say that the agent forms the belief P^* with which the probability P is selected. Given this we would say that the agent believes that the dynamical system is $(\Omega, \mathcal{F}, P, T)$.

It is important to stress that the above selection procedure does not change with additional data. This is the case since P^* itself is not just a prior probability subject to updating; it is already the limit of the updated beliefs given all the wealth of data we have provided the agent in the first place. Given the stability of P^* we can then proceed to specify the basic rationality axioms relative to the selection of P^*. Having done so we shall prove that the integral in Eq. (10) is well defined and hence a probability $P \in \mathcal{P}(\Omega)$ is, in fact, chosen by the agent. All our development will then focus on the characterization of P.

We now introduce an important result regarding the process of forming beliefs. We use the following notation

$$x^t = (x_t, x_{t+1}, x_{t+2}, \ldots)$$
$$x_{(t)} = (x_0, x_1, \ldots, x_{t-1})$$
$$\mathcal{F}^t = \sigma(x^t) \quad .$$

$\mathcal{F}^t$ is the σ-field of "future" events at date t. We write the conditional probability of a future event $S \in \mathcal{F}^t$ given the past history in the form

$$P^t(S \mid x_{(t)}) \quad .$$

Now recall that our context is that t is assumed large. This implies that a long history is available for learning and therefore the limiting behavior of the measure $P^t(\cdot|x_{(t)})$ should be of interest. This motivates

Definition 6: Regular conditional probabilities $P^t(\cdot|x_{(t)})$ and $Q^t(\cdot|x_{(t)})$ on (Ω,F) are said to agree for Q almost all histories if

$$\lim_{t\to\infty} \operatorname{Sup}_{S\epsilon F^t}|P^t(S|x_{(t)}) - Q^t(S|x_{(t)})| = 0 \qquad Q \text{ a.e.}$$

and we write $P^t \approx Q^t$ Q a.e. If they agree Q a.e. and P a.e. then we write $P^t \approx Q^t$ Q a.e, P a.e.

Recall now a few definitions from measure theory. A probability measure P is said to be absolutely continuous with respect to Q (denoted by $P \ll Q$) if $Q(S) = 0$ implies $P(S) = 0$ for $S \epsilon F$. The two probabilities are said to be equivalent if $P \ll Q$ and $Q \ll P$. The two probabilities are said to be singular (denoted by $P \perp Q$) if there exist measurable sets A and B such that $A \cap B = \emptyset$ $A \cup B = \Omega$, $P(B) = 0$ and $Q(B) = 1$.

We can now state the following result:

Theorem 3: (Blackwell and Dubins [1962]). Suppose that P and Q are probability measures on (Ω,F) and $Q \ll P$. Then for each t and for every regular conditional probability P^t of the future given the past there exists a corresponding conditional probability Q^t such that

$$P^t \approx Q^t \qquad Q \text{ a.e.}$$

We remark that the technical issue of selecting regular conditional probabilities P^t and Q^t should be entirely disregarded here since

$\Omega = X^{\infty}$ is a complete and separable metric space (see Blackwell and Dubins [1975] and Ash [1972] page 265). The significance of Theorem 3 will be discussed below.

5.2 Axioms of Rationality.

Before formulating our two axioms we state the following:

Definition 7: We say that an agent's probability $Q \in \mathcal{P}(\Omega)$ is compatible with the data if

(a) $(\Omega, \mathcal{F}, Q, T)$ is stable with a stationary measure m. That is, for all $S \in \mathcal{F}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q(T^{-k}S) = m(S)$$

(b) Q satisfies the agent-specific tightness condition on Π (Assumption 4, Section 2). That is, for every ϵ there is a compact set $K_\epsilon \subset \Omega$ such that $Q(K_\epsilon) > 1 - \epsilon$.

We define now the agent's acceptable set $B(\Pi)$

(11a) $\quad B(\Pi) = \{Q \in \mathcal{P}(\Omega): Q$ is compatible with the data$\}$.

Also, relative to any measurable Set $S \in \mathcal{F}$ define

(11b) $$B_S = \{Q \in B(\Pi): \ Q(S) > 0\}.$$

Keeping in mind that the sets $B(\Pi)$ and B_S are agent specific we turn now to the axioms on the selection of P^*.

Axion 1 (Compatibility with the data): An agents forms a belief P^* with $B(\Pi)$ as its support.

<u>Axiom 2</u> (Continuity with respect to the data): If for $S \in F$ $m(S) > 0$ then $P^*(B_S) > 0$.

<u>Discussion</u>. Axiom 1 is the crucial axiom of rationality. It requires of any agent to form belief P^* which places probability 1 on $B(\Pi)$; that is, on the set of probabilities in the agent specific set of acceptable probabilities. A rational agent is not permitted to put positive measure on probabilities which would be contradicted by the data or that would conflict with his own tightness conditions.

Axiom 2 states that if an agent knows that the long run frequency of a set S is $m(S) > 0$ then he will place positive P^* measure on those Q in $B(\Pi)$ with $Q(S) > 0$. To see the meaning of Axiom 2 note that Axiom 1 says that if $m(S) > 0$ the agent must assign his P^* measure on $Q \in P(\Omega)$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q(T^{-k}S) = m(S) > 0.$$

This last condition requires $Q(T^{-k}S) > 0$ for most large k. If the agent knew that the system is stationary he would choose only among Q such that $Q(S) = Q(T^{-k}S)$ for $k \geq 0$. This would then <u>prove</u> that $Q(S) > 0$ and Axiom 2 would not be needed. Without the knowledge of stationarity it is conceivable that the agent selects Q with $Q(S) = 0$ but $Q(T^{-k}S) > 0$ for large k. Given the fact that agents do not know if the true measure is stationary, Axiom 2 proposes that when an agent observes $m(S) > 0$ he will place <u>some</u> positive P^* measure on $Q \in B(\Pi)$ with $Q(S) > 0$.

We think of Axiom 2 as a continuity axiom. This is so since when $m(S) > 0$ the agent is <u>certain</u> that for k large enough future S sets (i.e. $T^{-k}S$) must have positive probability. This means that he must place positive P^* measure on those $Q \in B(\Pi)$ with $Q(T^{-k}S) > 0$ for k large enough. What Axiom 2 says is that if the agent is certain that $(T^{-k}S)$ have positive probability for large k then he cannot assign 0 measure to probabilities Q with $Q(S) > 0$.

Returning now to Theorem 3 suppose that two agents believe that P and Q (respectively) are the true probabilities with $P \neq Q$ but that P and Q are equivalent. Since P and Q agree only on the null sets, the numerical probability assessments of the two agents may be drastically different. In fact, it is standard in the economics literature to assume the equivalence of subjective probabilities whenever heterogeneity of beliefs is introduced (see, for example, the important paper of Harrison and Kreps [1979] where this assumption is crucial for the validity of the conclusions). In our context ample past data is available and therefore it follows from Theorem 3 that, although $P \neq Q$, no significant difference of opinions will exist in their conditional probability assessment of future events given the past.

Now consider the specific context of our analysis where our agent knows m, the stationary measure of $(\Omega, \mathcal{F}, \Pi, T)$. We shall see below that the process of forming a belief by an agent involves a probability measure Q which is both stationary as well as equivalent to m. It follows from Theorem 3 that Q and m have the same conditional probabilities of future events given the past. This means that from the economic point of view of a rational agent in our context, there is no difference between Q and m. In the formal development below we shall, however, examine conditions under which $\overline{Q} = m$.

5.3 The Main Theorem

We now state and prove the main conclusion of this paper.

MAIN THEOREM: Given a dynamical system $(\Omega, \mathcal{F}, \Pi, T)$ let an agent form a belief P^* which satisfies Axioms 1 and 2. Then the integral

$$P = \int_{\mathcal{P}(\Omega)} \mu P^*(d\mu)$$

is well defined and $P \in B(\Pi)$. Moreover, there exist probabilities P_a and P_0 on $(\Omega, \mathcal{F})$ and a constant $0 < \lambda_p \leq 1$ such that

(i) P has a unique representation

$$P = \lambda_p P_a + (1-\lambda_p) P_o \tag{12a}$$

where P_a and m are equivalent while P_0 and m are singular

$$m(A) = 1 \text{ and } P_0(B) = 1.$$

(ii) $(\Omega, \mathcal{F}, P_a, T)$ and $(\Omega, \mathcal{F}, P_0, T)$ are stable with stationary measures $\overline{P}_a$ and $\overline{P}_0$ such that for all $S \in \mathcal{F}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P_a(T^{-k}S) = \overline{P}_a(S) \tag{12b}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P_0(T^{-k}S) = \overline{P}_0(S) \tag{12c}$$

where

$\overline{P}_a$ and m are equivalent

$\overline{P}_0 \ll m$.

If (Ω, F, Π, T) is ergodic and $0 < \lambda_p < 1$ then we have

$$\overline{P}_a = \overline{P}_0 = m .$$

(iii) There exist regular versions of the conditional probabilities P^t, m^t and P_0^t and densities

$$\psi_m = \frac{dm}{dP} \qquad \psi_0 = \frac{dP_0}{dP}$$

which satisfy m a.e. for $S \in F^t$

(12d) $$P^t(S|x_{(t)}) \approx \lambda_p m^t(S|x_{(t)})\psi_m(x_{(t)}) + (1-\lambda_p)P_0^t(S|x_{(t)})\psi_0(x_{(t)}) \quad .$$

<u>Proof of Main Theorem</u>

We start by demonstrating the existence of the choice P. Since $\Omega = X^\infty$ with $X \subset \Re^N$ is a complete and separable metric space, the space, $P(\Omega)$ endowed with the topology of weak convergence is a complete and separable metric space (see Parthasarathy [1967], chapter II.6). Hence if we let $B(P(\Omega))$ be the σ-field of the Borel subsets of $P(\Omega)$ then the space $(P(\Omega), B(P(\Omega)))$ is a Borel space. Denote by P^* any probability on the space $(P(\Omega), B(P(\Omega)))$.

Now consider the real valued function $f:P(\Omega) \times F \to \Re$ defined by

$$f(\mu, S) = \mu(S) \qquad \mu \in P(\Pi), \qquad S \in F.$$

For any P^* which satisfies Axiom 1 we have

(13) $$\int_{P(\Omega)} f(\mu,S)P^*(d\mu) = \int_{B(\Pi)} f(\mu,S)\ P^*(d\mu).$$

if the integral is defined. From Assumption 4 about the tightness conditions it follows that the family $B(\Pi)$ is tight and hence $B(\Pi)$ is compact (see Billingsley [1968], Appendix III, Theorem 6). From the definition we note that $B(\Pi)$ is a convex set. Moreover, it follows from the Prohorov Theorem that on $B(\Pi)$ the topology of weak convergence and the topology generated by the Prohorov metric are equivalent. It then follows that on $B(\Pi)$ the function $f(\mu,S)$ is continuous in μ for each fixed $S \in F$. Denote by $\rho(P,Q)$ the Prohorov metric on $B(\Pi)$.

Now let $B(n) = \{B_1^n, B_2^n, \ldots, B_{m(n)}^n\}$ be a sequence of partitions of $B(\Pi)$ where $B(n)$ is a refinement of $B(n-1)$. Let

$$\gamma(B_i) = \sup_{\substack{P \in B_i \\ Q \in B_i}} \rho(P,Q)$$

and

$$\epsilon_n = \max_{1 \le i \le m(n)} \gamma(B_i^n).$$

Let the sequence of partitions $B(n)$ satisfy $\epsilon_n \downarrow 0$. By the continuity of f we can obtain the integral in (13) as the following limit

$$\int_{B(\Pi)} f(\mu,S)P^*(d\mu) = \lim_{n\to\infty} \sum_{i=0}^{m(n)} f(\mu_i^n,S)P^*(B_i^n) \qquad \mu_i^n \in B_i^n$$

independent of the selection $\mu_i^n \in B_i^n$ and independent of the sequence $B(n)$. Note that for each $S \in F$

$$\sum_{i=0}^{m(n)} f(\mu_i^n,S)P^*(B_i^n) = \sum_{i=1}^{m(n)} \mu_i^n(S)P^*(B_i^n) = P^n(S).$$

Since $\sum_{i=1}^{m(n)} P^*(B_i^n) = 1$, $\mu_i^n \in B(\Pi)$ and $B(\Pi)$ is convex, the set function $P^n(\cdot)$ is a measure in $B(\Pi)$. However $B(\Pi)$ is compact hence P^n converges weakly to a measure $P \in B(\Pi)$. This proves that for every $S \in F$

$$P(S) = \int_{B(\Pi)} f(\mu,S)\ P^*(d,\mu)$$

is well defined and $P \in B(\Pi)$

Next we consider the decomposition of P. It follows from the Borel decomposition Theorem (See Royden [1988], page 278) that there exists probabilities P_a and P_o on (Ω,F), sets $A \subset \Omega$ and $B = \Omega - A$, and a constant $0 \leq \lambda_p \leq 1$ such that

$$P = \lambda_p P_a + (1-\lambda_p)P_o \tag{14}$$

where $P_a \ll m$, $P_o \perp m$, $\lambda_p = P(A)$, $m(A) = 1$ and $m(B) = 0$. For any set $S \in F$, if $P(A) > 0$ and $P(B) > 0$ we have

$$P_a(S) = \frac{P(A \cap S)}{P(A)}$$

$$P_0(S) = \frac{P(B \cap S)}{P(B)} .$$

By Axiom 2 $m(A) = 1$ implies $P(A) > 0$ and hence $0 < \lambda_p \leq 1$. We do not exclude $\lambda_p = 1$ and $P(B) = 0$.

We shall now show that P_a and m are equivalent. From the Borel decomposition theorem we already have

$$P_a \ll m.$$

To prove that $m \ll P_a$ suppose that it is false. Thus let $S \in F$ and $P_a(S) = 0$ while $m(S) > 0$. From Axiom 2 it follows that $P(S) > 0$. But then we have

$$0 < P(S) = \lambda_p P_a(S) + (1-\lambda_p)P_0(S) = (1-\lambda_p)P_0(S)$$
$$0 < m(S) = m(S \cap A) + m(S \cap B) = m(S \cap A).$$

Hence $m(S \cap A) > 0$ and $P_0(S) > 0$. Now consider $\hat{S} = S \cap A$. Clearly $m(\hat{S}) > 0$ but $\hat{S} \subset S$ implies $P_a(\hat{S}) = 0$ and $\hat{S} \subset A$ implies $P_0(\hat{S}) = 0$ (P_0 and m are singular). Hence $P(\hat{S}) = 0$ and this contradicts Axiom 2.

We now demonstrate that (Ω, F, P_a, T) and (Ω, F, P_0, T) are stable with stationary measures $\overline{P}_a$ and $\overline{P}_0$ and that $\overline{P}_a$ and m are equivalent. For any set $S \in F$ we have that

$$\frac{1}{n}\sum_{k=0}^{n-1} P(T^{-k}S) = \lambda_p \frac{1}{n}\sum_{k=0}^{n-1} P_a(T^{-k}S) + (1-\lambda_p)\frac{1}{n}\sum_{k=0}^{n-1} P_0(T^{-k}S) \ . \tag{15}$$

Since $P \in B(\Pi)$ the left hand side of (15) converges to $m(S)$ for all $S \in F$. Now since $P_a \ll m$ and since m is stationary it follows from Theorem 2 of Gray and Kieffer [1980] that P_a is asymptotically mean stationary and hence for all $S \in F$ the limit

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} P_a(T^{-k}S) = \overline{P}_a(S) \quad \text{exists} \ . \tag{15'}$$

Combining (15) and (15′) leads to the conclusion that for all $S \in F$ the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P_0(T^{-k}S) = \overline{P}_0(S) \quad \text{also exists}$$

and that for all $S \in F$

(16) $$m(S) = \lambda_p \overline{P}_a(S) + (1-\lambda_p)\overline{P}_0(S) \quad .$$

This shows that both (Ω,F,P_a,T) and (Ω,F,P_0,T) are stable dynamical systems with stationary measures $\overline{P}_a$ and $\overline{P}_0$. It is immediate from (16) that $\overline{P}_0 \ll m$. We need to prove that $\overline{P}_a$ and m are equivalent.

To show that $\overline{P}_a$ is equivalent to m it is immediate from (16) that $m(S) = 0 \Rightarrow \overline{P}_a(S) = 0$ hence $\overline{P}_a \ll m$. To prove that $m \ll \overline{P}_a$ assume the contrary and select $S \in F$ with $\overline{P}_a(S) = 0$ but $m(S) > 0$. Since P_a is equivalent to m, $P_a(S) > 0$. Now define

$$\hat{S} = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} (T^{-k}S)$$

$$= \limsup_{k\to\infty} (T^{-k}S) \quad .$$

We claim that $\overline{P}_a(\hat{S}) = 0$. This is so since

$$\overline{P}_a(\hat{S}) = \lim_{n\to\infty} \overline{P}_a\Big(\bigcup_{k=n}^{\infty} T^{-k}S\Big) \leqq \overline{P}_a\Big(\bigcup_{k=1}^{\infty} T^{-k}S\Big) \leqq \sum_{k=1}^{\infty} \overline{P}_a(T^{-k}S) = 0 \quad .$$

But $\hat{S}$ is an invariant set so that on $\hat{S}$, $\overline{P}_a(\hat{S}) = P_a(\hat{S}) = 0$. Hence $m(\hat{S}) = 0$. By Fatou's Lemma

$$m(\hat{S}) = m(\limsup_{k\to\infty} T^{-k}S) \geqq \limsup_{k\to\infty} m(T^{-k}S) = m(S) > 0$$

and this is a contradiction hence $m \ll \overline{P}_a$. This concludes the proof that $\overline{P}_a$ and m are equivalent.

To prove that in the ergodic case

$$\overline{P}_a = P_0 = m$$

it is sufficient to prove that $P_a = m$ since the conclusion follows from (16) and $0 < \lambda_p < 1$. To prove $P_a = m$ recall that $P_a \ll m$ hence there exists an m-integrable function g such that for all $S \in F$

$$P_a(S) = \int_S g(\omega)m(d\omega) .$$

Hence

$$P_a(T^{-k}S) = \int_{T^{-k}S} g(\omega)m(d\omega)$$

By the change of variables theorem and the stationarity of m we have that

$$\int_{T^{-k}S} g(\omega)m(d\omega) = \int_S g(T^k\omega)(mT^{-k})(d\omega) = \int_S g(T^k\omega)m(d\omega) .$$

It then follows that

$$\overline{P}_a(S) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P_a(T^{-k}S) = \int_S (\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^{-k}\omega))m(d\omega) .$$

Now use Theorem 2(iv) to conclude that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^{-k}\omega) = E_m g = 1 .$$

Hence

$$\overline{P}_a(S) = m(S) \qquad \text{all} \quad S \in F \ .$$

We finally turn to the conditional probabilities. Since Ω is a complete and separable metric space it follows from Theorem 3 that for all $S \in F^t$

$$P_a^t(S|x_{(t)}) \approx m^t(S|x_{(t)}) \quad \text{a.e.} \quad m \quad \text{and} \quad P_a \quad .$$

A standard argument leads to

$$P^t = \lambda_p P_a^t \psi_m + (1-\lambda_p) P_0^t \psi_0$$

where $\psi_a = \frac{dP_a}{dP}$ and $\psi_0 = \frac{dP_0}{dP}$. We conclude that

$$P^t \approx \lambda_p m^t \psi_m + (1-\lambda_p) P_0^t \psi_0 \quad \text{a.e.} \quad m \quad \text{and} \quad P_a \quad . \qquad \blacksquare$$

5.4 Some Implications of the Main Theorem and the Structure of Non-Stationarity

In this section we explore some implications of the Main Theorem with a view to clarify the nature of non-stationary beliefs.

Proposition 1. (Time discounting under non-stationarity): Let $S \in F$. If $m(S) = 0$ then $\lim_{k\to\infty} P(T^{-k}S) = 0$.

Proof: See Rechard [1956] and Theorem 5 of Gray and Kieffer [1980]. $\blacksquare$

For an intuitive sense note that since $P \in B(\Pi)$ we have that for all $S \in F$ the limit

$$\lim_{k\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P(T^{-k}S) = m(S) \quad \text{exists} \quad .$$

Hence, if $m(S) = 0$ then the sequence $P(T^{-k}S)$ itself must converge to 0. To provide an economic interpretation of Proposition 1 recall that if $K = T^{-n}S$ then for each $x \in K$, $T^n x \in S$: in n steps K will be transformed into S. For this reason we suggested that one may think of K as a future set S, n dates into the future. This interpretation is particularly applicable to shift transformations. When $m(S) = 0$ the agent knows that the long run frequency of S is equal to zero thus a belief that $P(S) > 0$ must be due to <u>his belief in non-stationarity</u>. In this case the condition of compatibility with the data requires him to believe that the probability that S occurs in the future <u>declines</u> to 0.

The validity of our interpretation which stresses non-stationarity can be formally seen as follows:

$$P(T^{-k}S) = \lambda_p P_a(T^{-k}S) + (1-\lambda_p)P_0(T^{-k}S) \quad .$$

If $m(S) = 0$ then the invariance of m implies $m(T^{-k}S) = 0$. However, since $P_a \ll m$ it follows that $P_a(T^{-k}S) = 0$. We can thus conclude that

$$P(T^{-k}S) = (1-\lambda_p)P_0(T^{-k}S)$$

and proposition 1 implies that $P_o(T^{-k}S) \to 0$. The point to be made is that the essential non-stationarity of P is to be found in the measure P_0.

The discussion above suggests that we may gain additional insight by considering the celebrated theory of recurrence. A few definitions will set the stage for our discussion.

Definition 8: Let $S \in F$. A point $x \in S$ is said to be recurrent with respect to T if there is a finite integer $k(S) \geq 1$ such that $T^{k(S)}x \in S$. A set $S \in F$ is said to be recurrent if almost every point in S is recurrent. A dynamical system is recurrent if every set in F is a recurrent set.

In a recurrent dynamical system we can define, for each $S \in F$, the set

$$\hat{S} = \bigcup_{k=1}^{\infty} T^{-k}S \quad .$$

$\hat{S}$ is the set of all points which enter S in one or more steps. The set $N(S)$ defined by

$$N(S) = S - \hat{S}$$

is then the set of members of S which do not return to S. In a recurrent system $P(N(S)) = 0$ for all $S \in F$. It is then clear that in a recurrent system all sets of positive measure recur infinitely many times. The celebrated Poincaré Recurrence Theorem (1899) state that any stationary dynamical system is recurrent. although we cannot use this theorem we could identify recurrence in our system by thinking of it as consisting of two distinct dynamical systems (Ω, F, P_a, T) and (Ω, F, P_0, T). Our Main Theorem proves that both systems are stable with m as their common stationary measure. There is, however, one crucial difference between these two systems which we shall explore: P_a and m are equivalent whereas P_0 and m are singular.

Proposition 2: The dynamical system (Ω, F, P_a, T) is recurrent.

Proof: Since $P_a \ll m$ and $m \ll P_a$ we can use Theorem 6.4.3 of Gray [1988] which asserts that for a stable system (Ω, F, P_a, T) the condition $P_a \ll m$ and $m \ll P_a$ is equivalent to the property of recurrence. ■

What Proposition 2 reveals is that although the dynamical system (Ω, F, P_a, T) is not stationary it acts like a stationary system. This highlights again our point made earlier that the non-stationarity of P is due mostly to the P_0 component.

Definition 9: A set W is a wandering set if for all $k = 1,2,3,\ldots$ $W \cap (T^{-k}W) = \emptyset$ hence the sequence of sets $\{(T^{-k}W); k = 1,2,\ldots\}$ is pairwise disjoint.

Note first that if for $S \in F$ we consider the set $N(S)$ in (16) we find that if $x \in N(S)$ then $T^k x$ does not return to S. This means that $N(S) \cap T^{-k}N(S) = \emptyset$ for $k = 1,2,\ldots$ and hence $N(S)$ is a wandering set. What Proposition 2 says is that $P_a(W) = 0$ for all wandering sets W in Ω.

We now turn our attention to (Ω, F, P_0, T).

Proposition 3: There exists wandering sets $W \subset \Omega$ with $P_0(W) > 0$.

Proof: We need to show that the dynamical system (Ω, F, P_0, T) is not recurrent. Suppose it is. Since it follows from our Main Theorem that (Ω, F, P_0, T) is stable with m as its stationary measure it follows from Theorem 6.4.3 of Gray [1988] (stated above in the proof of Proposition 2) that P_0 is equivalent to m. This contradicts the conclusion of the Main Theorem which states that P_0 and m are singular. ■

We have arrived at two interesting characterizations of the non-stationary measure P_0:

(a) for the set B specified in the Main Theorem we have $P_0(B) = 1$. However, any subset $S \subset B$ is <u>dissipative</u> in the sense that

$$\lim_{k\to\infty} P_0(T^{-k}S) = 0$$

(b) there exist wandering sets W with $P_0(W) > 0$.

In Proposition 1 we have suggested that dissipative sets can be thought of as giving rise to discounted probability of recurrence due to non-stationarity. In addition we may also note that the sets A and B satisfy

$$P_a(A) = 1 \qquad P_0(A) = 0$$
$$P_a(B) = 0 \qquad P_0(B) = 1 \ .$$

Turning to the second characterization of non-stationarity we have stressed our interest in non-stationary processes as a reflection of the complexity of the process of structural change. This includes technical innovations, the creation of new processes, new products, new organizational structures etc. The dynamic transformation of a wandering set as a representation of the process of structural change is very interesting. It simply says that the sequence of structural changes is a sequence of events <u>each of which needs to be thought of as different from anything that happened in the past</u>. These events have zero probability under the stationary measure hence $m(W) = m(T^{-k}W) = 0$ for all $k \geq 1$. However, under the non-stationary measure

$$P_0(W) > 0 \quad \text{and} \quad P_0(W) \neq P_0(T^{-k}W).$$

We insisted earlier that T is not invertible. We now clarify this point.

<u>Proposition 4</u>: Let the dynamical system (Ω, F, Π, T) be stable. If T is invertible then $\Pi \ll m$ and consequently $P = P_a$ and $\lambda_p = 1$.

<u>Proof</u>: If T is invertible it follows from corollary 6.3.2 of Gray [1988] that $\Pi \ll m$. Since this is known to agents it follows that $P \ll m$ as well. Recalling that $P(S) = \lambda_p P_a(S) + (1-\lambda_p)P_0(S)$ for all $S \in F$, let $S = B$ where $P_a(B) = 0$ and $P_0(B) = 1$. This implies $P(B) = 1 - \lambda_p$. However, since $P \ll m$ and by the Main Theorem $m \ll P_a$, the condition $P_a(B) = 0$ implies $P(B) = 0$ and hence $\lambda_p = 1$. ∎

Proposition 4 reveals that invertible T render stable systems essentially stationary.

6. <u>Discussion of the Results and Examples</u>.

The theory developed in this paper proposes that in a non-stationary environment the rationality of a belief should be judged only on the basis of its compatibility with the long-term, average patterns of the data generated by the stochastic dynamical system. The characterization of the Main Theorem shows that two intelligent agents knowing the same stationary measure and having the same, and extensive, amount of past data may end up with different probability beliefs. It is important to keep in mind that the agents know that in forming their beliefs they have used all available information and therefore their disagreements cannot be resolved by employing past data. Agents also know that the true Π is not known by anyone and is not knowable if past data is the only source to be consulated. This is what we called in the Introduction the "scientific gaps": they result in a speculative search for ideas about the structural causes of the random mechanism which generates the data. This search induces the formation of hypotheses, conjectures or theories and these are the main source of heterogeneity of beliefs.

Our conclusions are diametrically opposed to the "common prior" assumption advocated by most Bayesians. We think the reason for this discrepancy can be found in our basic position that one must think of the random nature of economic fluctuations as structurally caused and subject to further understanding as human knowledge improves. From a Bayesian view-point the formation of beliefs is only an expressions of the uncertainty of the agent. From our view-point the formation of beliefs $P^* \in \mathcal{P}(\mathcal{P}(\Omega))$ incorporates both the uncertainty of the agent given the data which he observes as well as his best hypotheses about the structural mechanism which generated the data. It is ultimately our ignorance which is the cause of diversity of beliefs.

We started this paper with the prototype example of the present value of profits, model (1)

$$P_t^* = \sum_{k=0}^{\infty} \gamma^{k+1} y_{t+k} .$$

We asked how would a rational agent form beliefs to evaluate the random prospect P_t^*. We can now use of Main Theorem to answer this question. Our theorem says that a rational agent will employ two probabilities: the stationary probability m and the non-stationary stable probability P_0

which has a stationary measure $\overline{P}_0$ such that $\overline{P}_0 \ll m$. These will be formed in accordance with our theory and their formation will be compatible with the data available. The agent will now evaluate p_t^* separately with respect to the conditional probabilities of m and P_0 to obtain

$$v_t^m(x_{(t)}) = \sum_{k=0}^{\infty} \gamma^{k+1} E_m(y_{t+k} \mid x_{(t)})$$

$$v_t^0(x_{(t)}) = \sum_{k=0}^{\infty} \gamma^{k+1} E_{p_0}(y_{t+k} \mid x_{(t)}) \ .$$

Together with m and P_0 the agent also selected λ_p to represent the probability which he assigns to the event that the process $\{y_t, \ t=0,1,2,\ldots\}$ is stationary. The agent's final conditional valuation is

$$E_p(p_t^* \mid x_{(t)}) \equiv v_t(x_{(t)}) = \lambda_p v_t^m(x_{(t)}) + (1 - \lambda_p) v_t^0(x_{(t)}) \ .$$

Having selected a distribution for p_t^* all other moments can also be calculated. For example the conditional variance of $v_t(x_{(t)})$ with respect to $x_{(t)}$ is calculated to yield

$$\begin{aligned} \mathrm{Var}(v_t(x_{(t)}) = \lambda_p^2 \mathrm{Var}(v_t^m(x_{(t)}) + (1 - \lambda_p)^2 \mathrm{Var}(v_t^0(x_{(t)}) \\ + 2\lambda_p(1 - \lambda_p)\mathrm{Cov}(v_t^m(x_{(t)}), v_t^0(x_{(t)}) \ . \end{aligned}$$

This last expression has some interesting implications to the debate on volatility of stock prices.

It is worth noting that since the stationary measure m is a probability on (X^∞, F), the agent takes as known all the normal and regular patterns of sequences of profits. To this extent the agent supports his belief with any chart, curve or other empirical pattern often used to study past data. One needs to note, however, that we are not talking here about historical charts or other patterns of stock prices. Rather, we are discussing the formation of belief about the dynamics of the profit sequence $\{y_t, t=0,1,2...\}$. Moreover, our theory does not imply any particular theory of equilibrium asset valuation. We certainly are proposing how agents should form rational beliefs about sequences of future profits, and this would be an essential ingredient in an equilibrium theory of asset valuation.

Economic applications of our approach will be developed in other papers. We shall now present two examples to assist the reader in understanding the ideas of the paper.

Example 2: Coin Tossing

Consider the sequence of independent random variables x_t. Under the true measure $\Pi = m$

$$x_t = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases} .$$

Since each infinite sequence $x = (x_0, x_1, x_2, ...)$ is a binary expansion of a real number in $[0,1]$ the stationary measure $\Pi = m$ is isomorphic to the Lesbegue measure on $[0,1]$.

Now consider the following belief P_y. Let $\{\gamma_t\}_{t=0}^{\infty}$ be a sequence of

small positive number $\gamma_t = \frac{1}{2(t+2)}$ $t = 0,1,2,...$. Now let the sequence $\{y_t\}_{t=0}^{\infty}$ be defined by

$$y_t = \begin{cases} 1 & \text{with probability } 1/2 + \gamma_t \\ 0 & \text{with probability } 1 - (1/2 + \gamma_t) \end{cases} .$$

Let P_y be the associated measure on infinite sequences y. It is clear that P_y and Π are equivalent and P_y is a rational belief.

Now consider a belief P_z which would have drastically different properties. Let $D = \{t_1, t_2, t_3, \ldots\}$ be an infinite sequence of "remote" dates such that

$$t_1 \geqq 2$$

$$t_n \geqq 2t_{n-1} \quad .$$

Now define the random variables z_t under P_z:

<u>If $t \notin D$</u>

$$z_t = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases} \quad .$$

<u>If $t \in D$</u>

$$z_t = \begin{cases} 1 & \text{with probability } 1/3 \\ 0 & \text{with probability } 2/3 \end{cases} \quad .$$

Note that for any cylinder set S the future set $(T^{-k}S)$ will fall on fewer and fewer dates in D and since these distances are more than geometric

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P_z(T^{-k}S) = m(S) \quad .$$

It is then clear that (Ω, F, P_z, T) is stable with m as its stationary measure. Nevertheless we shall now indicate that P_z and m are singular. To see this consider the following B set:

$$B = \left\{ x: \begin{array}{l} \text{the frequency of "1" at the infinite dates in} \\ D \text{ is } 1/3. \end{array} \right\}$$

By the strong law of large numbers we have

$$m(B) = 0$$

$$P_z(B) = 1 \quad .$$

On the other hand consider the A set

$$A = \left\{ x: \begin{array}{l} \text{the frequency of "1" in the sequence } (x_0, x_1, x_2, \ldots) \\ \text{is } 1/2 \text{ and this same frequency among the infinite} \\ \text{number of dates in } D \quad . \end{array} \right\}$$

Here we must have

$$m(A) = 1$$

$$P_z(A) = 0 \quad .$$

It then follows that in the representation

$$P_z = \lambda_p P_a + (1-\lambda_p) P_0$$

we have $\lambda_p = 0$ and $P_z = P_0$ and this violates Axion 2. To see why note that at date t we have $t_n \leqq t < t_{n+1}$ and the agent has n observations $(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$. An examination of the frequency of "1" among these dates cannot guarantee that he is either right or wrong about the probability of "1" at future dates $t_j \in D$ for $j > n$. Axiom 2 implies that he should give <u>some positive weight</u> to the <u>possibility</u> that the true probability is, in fact, m. Thus, any belief Q of the form

$$Q = \lambda_p m + (1-\lambda_p)P_z$$

with $\lambda_p > 0$ is compatible with the Main Theorem.

We can modify example 2 in a somewhat revealing manner. Suppose that $P_z = \Pi$ is the true non-stationary probability so that in fact it is true that at the remote dates in the set D the probability switches to 1/3. An agent who adopts the belief that m is the true measure is, in fact, selecting a belief Q such that

$$Q = \lambda_q m + (1-\lambda_q)Q_0$$

with $\lambda_q = 1$, some Q_0 and $Q = m$. Such a belief is compatible with the Main Theorem. But now suppose that an agent selects the belief $Q = \Pi$. Hence his belief is

(17) $$Q = \lambda_q m + (1-\lambda_q)\Pi$$

where Π and m are singular, $\lambda_q = 0$ thus $Q = \Pi$ and Π is the truth! Our theory proposes that this is not rational belief although the agent selected the truth as his belief!! This appears paradoxical. On a purely formal basis Axiom 2 requires the selection of $\lambda_q > 0$ in (17). The agent who believes that $Q = \Pi$ is then permitted to select λ_q as small as he pleases. In any decision-theoretic context all decisions under small λ_q will be the same as under $Q = \Pi$. The appearence of a paradox arises because the agent has absolutely no systematic way of selecting the set of true infinite dates D at which the probability switches. The agent has no knowledge that there exists a set D of remote dates at which the probability switches and even if he suspected that such is the fact and selected a set D at random, the probability of selecting the correct set is zero. Axiom 2 proposes that since the agent recognizes his ignorance he should place <u>some</u> positive probability on the possibility that the actually observed frequencies over all the dates represent the true probability.

<u>Example 3</u>: <u>Path Dependency</u>

Let $\{x_0, \epsilon_1, \epsilon_2, \ldots\}$ be an infinite sequence of i.i.d. random variables where

$$x_0 \quad \text{or} \quad \epsilon_t = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases} .$$

Now let $\{x_t\}_{t=0}^{\infty}$ be a sequence of random variables defined by

$$x_0 \quad \text{as specified above}$$

$$x_t = \alpha_t x_0 + \epsilon_t \qquad\qquad t = 1, 2, \ldots$$

where $\{\alpha_1, \alpha_2, \alpha_3, \ldots\}$ is a deterministic sequence of numbers taking values $\alpha_1 = 1$ or $\alpha_t = 0$.

This is a model of extreme path dependency: the effect of the initial value taken by x_0 continues to linger on forever but this effect depends upon α_t. Note that for a constant α_t there exists an initial distribution for x_0 which would make the sequence stationary. The first question which we investigate is whether the dynamical system generating the x_t is stable.

To examine the stability question we calculate first $\Pi(x_t, \alpha_t)$ - the unconditional distribution of x_t which clearly depends upon α_t. A direct calculation reveals that this distribution is as follows:

$$\underline{\Pi(x_t,\alpha_t)}$$

$\underline{\alpha_t = 0}$

x_t	0	1	2
$\Pi(x_t,0)$	1/2	1/2	0

$\underline{\alpha_t = 1}$

x_t	0	1	2
$\Pi(x_t,1)$	1/4	1/2	1/4

.

Stability requires that the following limit exists:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \Pi(x_k,\alpha_k) = m(x) \quad \text{exists} .$$

In the above, the function $\Pi(x_t,\alpha_t)$ takes two values represented by the vectors $(1/2,1/2,0)$ and $(1/4,1/2,1/4)$ which are then averaged by the frequency of their occurence. These frequencies are exactly the frequencies by which the α_t take the values 0 and 1. Thus a sufficient condition for stability in this case is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \alpha_k = \alpha \quad \text{exists} .$$

In this case the stationary distribution of x_t is simply

$$m(x) = \begin{pmatrix} \frac{1}{4}\alpha + \frac{1}{2}(1-\alpha) \\ \frac{1}{2} \\ \frac{1}{4}\alpha \end{pmatrix} .$$

The requirement of stability demands the convergence of all the joint distributions as well. Thus if $\Pi(x_t, x_{t+1}, \alpha_t, \alpha_{t+1})$ is the joint distribution of x_t and x_{t+1} it can easily be calculated to be

$\underline{\Pi(x_t, x_{t+1}, \alpha_t, \alpha_{t+1})}$

$\underline{\alpha_t = 0 \quad , \quad \alpha_{t+1} = 0}$

x_{t+1} / x_t	0	1	2
0	1/4	1/4	0
1	1/4	1/4	0
2	0	0	0

$\underline{\alpha_t = 0 \quad , \quad \alpha_{t+1} = 1}$

x_{t+1} / x_t	0	1	2
0	1/8	2/8	1/8
1	1/8	2/8	1/8
2	0	0	0

$\underline{\alpha_t = 1 \quad , \quad \alpha_{t+1} = 0}$

x_{t+1} / x_t	0	1	2
0	1/8	1/8	0
1	2/8	2/8	0
2	1/8	1/8	0

$\underline{\alpha_t = 1 \quad , \quad \alpha_{t+1} = 1}$

x_{t+1} / x_t	0	1	2
0	1/8	1/8	1/8
1	1/8	3/8	1/8
2	0	1/8	1/8

Stability in this case requires that the averaging of the four matrix values of the distribution $\Pi(x_t, x_{t+1}, \alpha_t, \alpha_{t+1})$ converges. To express this condition define the four sets

$$K_{0,0} = \{0,0\} \qquad K_{0,1} = \{0,1\}$$
$$K_{1,0} = \{1,0\} \qquad K_{1,1} = \{1,1\} .$$

Now consider the sequence of pairs $\{(\alpha_t, \alpha_{t+1}), t=0,1,2,\ldots\}$. The condition for stability is that the four following limits exist:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{K_{0,0}}(\alpha_k, \alpha_{k+1}) = \alpha_{0,0}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{K_{0,1}}(\alpha_k, \alpha_{k+1}) = \alpha_{0,1}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{K_{1,0}}(\alpha_k, \alpha_{k+1}) = \alpha_{1,0}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{K_{1,1}}(\alpha_k, \alpha_{k+1}) = \alpha_{1,1}$$

In the special case when

$$\alpha_{0,0} = \alpha_{0,1} = \alpha_{1,0} = \alpha_{1,1} = \frac{1}{4}$$

the stationary joint distribution $m(x_t, x_{t+1})$ is

$m(x_t, x_{t+1}) =$

x_{t+1} / x_t	0	1	2
0	5/32	6/32	1/32
1	6/32	9/32	1/16
2	1/32	2/32	1/32

The general pattern may now be spelled out. Let $\{\alpha_t\}_{t=1}^{\infty}$ be a sequence of real numbers taking values in a "state" spece $K \subseteq \Re$. Consider the product space K^{∞} with its Borel sets $L_K = B(K^{\infty})$. Clearly (K^{∞}, L_K) is a measurable space. Let T be the shift transformation on K^{∞} so that if $\alpha = (\alpha_1, \alpha_2, \alpha_3, \ldots)$ then

$$T^k\alpha = (\alpha_{k+1}, \alpha_{k+2}, \alpha_{k+3}, \ldots)$$

<u>Definition 10</u>: A sequence $\{\alpha_t\}_{t=1}^{\infty}$ is said to be <u>stable</u> if for any $S \in L_K$ the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_S(T^k\alpha) = m_\alpha(S) \quad \text{exists.}$$

If a sequence α is stable then m_α is a probability on (K^∞, L_K). We call m_α the <u>limit measure of α</u>.

Applying this definition to our problem it is seen that if (Ω, F, Π, T) is the dynamical system generating x_t then it is stable if the infinite sequence $\{\alpha_t\}_{t=1}^{\infty}$ is a stable sequence. The stationary distribution of $(x_{\tau_1}, x_{\tau_2}, \ldots, x_{\tau_n})$ can then be calculated by first calculating the joint distribution $\Pi(x_{\tau_1}, \ldots, x_{\tau_n}, \alpha_{\tau_1}, \ldots, \alpha_{\tau_n})$ and then computing the expectations under the stationary measure m_α induced by α. That is

$$(18) \qquad m(x_{\tau_1}, \ldots, x_{\tau_n}) = \int_{K_\infty} \Pi(x_{\tau_1}, \ldots, x_{\tau_n}, \alpha_{\tau_1}, \ldots, \alpha_{\tau_n}) m_\alpha(d\alpha)$$

Turning now to the question of heterogenously forming beliefs about Π we can focus on the sequence α. A belief Q_β will entail the selection of a sequence $\beta = \{\beta_t\}_{t=1}^{\infty}$. However, for $Q_\beta \in B(\Pi)$ the sequence β must be stable with a limit measure m_β satisfying $m_\beta = m_\alpha$. Clearly, the set of such Q_β measures is very large. To construct rational beliefs in accordance with the Main Theorem we can simply take for such a β

$$P(\beta, \hat{\lambda}_p) = \hat{\lambda}_p m + (1-\hat{\lambda}_p) Q_\beta \ .$$

If Q_β is not singular with m we can decompose it into $Q_\beta = \mu Q_{\beta_a} + (1-\mu)Q_{\beta_0}$ and then we have

$$P(\beta,\lambda_p) = \lambda_p P_a + (1-\lambda_p)P_0$$

where

$$1-\lambda_p = (1-\hat{\lambda}_p)(1-\mu)$$

$$P_a = \frac{\hat{\lambda}_p}{\lambda_p} m + \frac{\mu(1-\hat{\lambda}_p)}{\lambda_p} Q_{\beta_a}$$

$$P_0 = Q_{\beta_0} .$$

To see that we can easily find measures Q_{β_0} which are stable yet singular with m consider an infinite set D of dates t_j $j = 0,1,2,\ldots$ such that

$$t_0 \geq 2$$

$$t_n \geq 2t_{n-1} \qquad n \geq 1 .$$

Since α is stable there exists a set D such that the following limit exists

$$\lim_{n\to\infty} \frac{1}{n} \sum_{n=0}^{n-1} \alpha_{t_j} = \alpha^*$$

Now select a sequence $\hat{\beta}$ such that

(i) $\quad Q_{\hat{\beta}} \in B(\Pi)$

(ii) $\quad \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \hat{\beta}_{t_j} = \hat{\beta} \qquad$ with $\hat{\beta} \neq \alpha^*$.

It is clear that such a sequence can be constructed from any stable sequence α without changing the limit measure of the sequence. But now we claim that $Q_{\hat{\beta}}$ and m are singular. Hence, for any $0 < \lambda_p < 1$ the measure

$$P = \lambda_p m + (1-\lambda_p)Q_{\hat{\beta}}$$

is a rational belief.

7. REFERNCES

Ash, R. B. [1972] *Real Analysis and Probability* Academic Press.

Aumann, R. J. [1976] "Agreeing to Disagree," *Annals of Statistics,* 4 1236-1239.

Aumann, R. J. [1987] "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometric*, 55, 1-18.

Blackwell, D., and L. Dubins, [1962] "Merging of Opinions with Increasing Information," *Annaals of Mathematical Statistics*, 33, 882-886.

Blackwell, D. and L. Dubins, [1975] "On Existence and Non-Existence of Proper, Regular, Conditional Distributions," *The Annals of Probability,* 3, 741-752.

Diaconis, P. and D. Freedman [1986], "On the Consistency of Bayes Estimates," *The Annals of Statistics,* 14(1), 1-26.

Dowker, Y. N. [1951], "Finite and Sigma-Finite Invariant Measures," *Annals of Mathematics*, 54, 595-608.

Dowker, Y. N. [1955], "On Measurable Transformations in Finite Measure Spaces," *Annals of Probability*, 8, 962-973.

Fontana, R. J., R. M. Gray and J. C. Kieffer [1981], "Asymptotically Mean Stationary Channaels," *IEEE Transactions on Information Theory,* 27. 308-316.

Freedman, D. [1963], "On the Asymptotic Bahavior of Bayes Estimates in the Discrete Case I," *Annals of Mathematical Statistics,* 34, 1386-1403.

Freedman, D.[1965], "On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II," *Annals of Mathematical Statistics*, 36, 454-456

Gray, R. M. and J. C. Kieffer [1980], "Asymptotically Mean Stationary Measures," *Annals of Probability,* 8, 962-973.

Gray, R. M. [1988], *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag, New York.

Harrison, M. and D. Kreps [1979], "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory*, 20, 381-408.

Kieffer, J. C. and M. Rahe [1981], "Markov Channels are Asymptotically Mean Stationary," *Siam Journal of Mathematical Analysis*, 12, 293-305.

Krylov, N. and N. Bogoliouboff [1937], "La Théorie Général de la Mesure Dans Son Application a L'étude de Systémes Dynamiques de la Mécanique non Linéaire" *Annals of Mathematics*, 38, 65-113.

Neveu, J. [1965], *Mathematical Foundations of the Calculus of Probability, Holden-Day, San Francisco, CA.*

Oxtoby, J. C. [1952], "Ergodic Sets," *Bull. of the American Math. Society*, 58, 116-136.

Parthasarathy, K. R. [1967], *Probability Measures on Metric Spaces*, Academic Press, New York and London.

Petersen, K. [1983], *Ergodic Theory*, Cambridge University Press, Cambridge, England.

Rechard, O. W. [1956], "Invariant Measures for Many-One Transformation," *Duke Journal of Mathematics*, 23, 477-488.

Royden, H. L. [1988], *Real Analysis*, Third Edition, MacMillan Publishing, New York and London.