JOURNAL OF ECONOMIC THEORY **74**, 266–296 (1997)
ARTICLE NO. ET962254

Convergence in Economic Models with Bayesian Hierarchies of Beliefs*

Yaw Nyarko

New York University, 269 Mercer St. Rm 723, New York, New York 10003

Received March 2, 1995; revised August 12, 1996

I study a model where hierarchies of beliefs (the beliefs about the beliefs of other agents, etc.) are important. I provide conditions under which optimal actions of agents will converge to the Nash equilibrium of the model characterized by the true, previously unknown "fundamentals." The conditions are (i) a contraction property on the best-response mappings and (ii) a mutual absolute continuity condition on beliefs. Violation of (i) may result in an "anything is possible" result: any stochastic process of actions is consistent with maximizing behavior and Bayesian updating. Violation of (ii) may result in cyclical behavior of actions on each sample path. *Journal of Economic Literature* Classification Numbers: C70, C73, D81, D82, D83, D84. © 1997 Academic Press

1. INTRODUCTION

For long economists have known that the beliefs of agents in an economy affect the outcomes in the economy. Not only that, it has also been recognized that the beliefs that agents have about the actions *and* beliefs of other agents are also important in determining the evolution of the economy. Keynes, in a famous passage in the General Theory, likened professional investment to a beauty contest where predicting the average opinions of others is essential. Almost any comment about investor behavior on the financial markets stress the importance of investors predictions of the "market"—in particular, investors predictions of the average opinions mentioned by Keynes.

Despite the obvious importance of the above to economics there has been a very small literature on this topic. The extreme reliance of economics on "equilibrium" (e.g., Nash equilibrium, common priors, etc.) where all

* I thank the C. V. Starr Center and the Presidential Fellowship Fund at New York University for their generosity. I also thank Professors Larry Blume, David Easley, Jim Jordan, and Mike Woodford for many helpful discussions. All errors in this paper are my own responsibility. I also thank an anonymous referee for his/her comments.

0022-0531/97 $25.00

agents have the same beliefs and know everything there is to know, rules out the phenomena mentioned above, *a priori*! There are many who believe that the "infinite regress" problem introduced by considering such "average opinions about the average opinions of others," makes these models intractable. This paper contributes to a very small literature that tries to explicitly model the issues mentioned earlier, and in particular to study economies where such beliefs about beliefs may be important and to show that such models may indeed be "tractable."

The contribution of this paper is twofold. First, a general formal model is provided where the issues mentioned above may be studied. This draws upon the work on hierarchies of beliefs of [1] and [20]. (See [23] for more on the formal model and further references to this literature.) Second, within the context of this fairly abstract mode, conditions are provided under which over time the optimal actions of agents will converge to the Nash equilibrium of the model under the true, previously unknown, "fundamentals."

One of the first papers to provide a convergence result in this literature is [31] who studies the model of a continuum of firms facing uncertainty about parameters of a fixed linear demand curve. Townsend proves that over time there will be convergence to Nash equilibrium of the model characterized by the true parameter values. [11] studies further the same model and obtains the convergence result under much weaker assumptions. In comparison to the work of this paper, the above very important papers should be considered as *examples*. In particular, both papers exploit critically the structure of the linear model of firm behavior to setup the model and obtain all results.

This paper begins with an abstract model of an economy where all agents have imperfect information about fundamentals in the economy and the beliefs and actions of the other agents. Two notions of a type are defined: A Savage–Bayesian type of an agent specifies that agent's utility parameters as well as that agent's beliefs about the actions and beliefs of all the other agents in the economy. This may be thought of as the most "comprehensive" notion of a type since it specifies "everything" required for decision-making. A second notion of a type is a Harsanyi type. This is a notion of a type that only specifies an agent's utility parameter and that agent's beliefs about the utility parameters of other agents. This is a sparser notion of a type since, among other things, it does not specify what that agent believes about the *actions* of others. (This is the notion of a type typically used in Game Theory, and in particular in [14], when talking about a Harsanyi Bayesian equilibrium; hence its name. See [24] for more on the various definitions of a "type." The name "Savage–Bayesian" is of course in reference to the axioms of [30], and the Bayesian updating these axioms imply.)

A contraction property on the best-response functions is identified. This property is easily checked from the "fundamentals" of the model. For example, in the model of firms studied in the papers of Townsend and Feldman, this contraction property holds whenever the slope of the demand curve is less than one in absolute value. Under this contraction property each agent's optimal action may be stated in terms of her "sparser" Harsanyi type. With this, a martingale argument may be used to conclude that the optimal actions of agents must converge to the Nash equilibrium of the model characterized by the true underlying previously unknown "fundamentals."

This paper is organized as follows: In Section 2, an "example" is presented, which is essentially the model studied by Townsend and Feldman. I sketch how the convergence result is obtained in the example in three "STEPS." In the main part of the paper I show how these STEPS are generalized in the abstract model which may be non-linear and multivariate (i.e., may have a vector choice variable.) In the Section 9 I indicate what could go wrong when the key assumptions are violated. I show that a violation of the contraction property may result in an "anything is possible" result where any stochastic process of actions consistent with maximizing behavior, common knowledge of it, and Bayesian updating. I show by an example that a violation of the mutual absolute continuity condition may result in cyclical behavior of actions on each sample path.

As regards the previous literature I have already mentioned the papers [31] and [11]. [7] and [27] also study the "example" of the model of firms. [5] and [17] also obtain results on the convergence of beliefs under different assumptions. These two papers do not model the hierarchies of beliefs of agents. Further, their results typically require that the space of types be finite or countable and will in general fail if the set of different possible types or beliefs of agents may be indexed by say the unit interval. (In particular, if the type space is in "canonical form" where each agent chooses a different play and each such play is pure (i.e., does not involve randomization), then the Kalai and Lehrer assumptions require the type space to be finite or countable. See [21] for details.).

The assumptions of this paper are more related to those of [15], [16] and [22]. However those papers exploit the special structure of repeated finite action games and would not work for the models studied in this paper. The work of this paper complements the recent work on non-Bayesian "learning," and in particular the literature on ordinary least-squares learning of [8], [12] and [19]. This paper is the "Bayesian Learning" equivalent of that literature. This paper is the multi-agent extension of the work on single-agent Bayesian learning models of began by [6] and others. This paper owes that literature an intellectual debt.

2. A MOTIVATING EXAMPLE

Suppose that there is a set of agents indexed by the unit interval $I=[0,1]$ and uniformly distributed along that interval. For technical reasons suppose also that there are finitely many classes of agents within that interval with all agents of the same class identical in all respects. Fix any date n. At that date agent i must choose an output level y_{in}. The aggregate output is then $y_n \equiv \int_0^1 y_{in}\, di$. The price of that output is determined via a linear demand curve $p_n=\alpha-\beta y_n+\varepsilon_n$, where α and β are fixed parameters, "the fundamentals," and ε_n is the date n shock to the demand curve—a zero mean unobserved random variable. Suppose that there is imperfect information over the parameter α. The value of the parameter β is common knowledge. The cost to firm i of choosing the output y_{in} is $c(y_{in})=0.5y_{in}^2$. The profit of firm i is then $p_n y_{in}-0.5y_{in}^2$. Let E_{in} denote the date n "expectations operator" of agent i. The profit maximizing output of firm i is then $y_{in}=E_{in}p_n=E_{in}\alpha-\beta E_{in}y_n$. Notice that to choose an optimal action agent i must form a belief over both the fundamentals, α, and the (aggregate) actions of other agents, y_n. Hence, *maximizing behavior* of firms (written as (MB)) implies the following:

$$\text{(MB)} \qquad y_{in}=E_{in}\alpha-\beta E_{in}y_n. \tag{2.1}$$

(The optimal output level should really be $y_{in}=\mathrm{Max}\{0, E_{in}\alpha-\beta E_{in}Y_n\}$ to ensure non-negative outputs of each agent. In this example this is ignored. In the formal part of the paper, it will be shown that non-negative outputs can easily be handled using the techniques developed here.)

Given any "random variable" x let $G_n x$ denote the "average opinion" of x, i.e., the average of the date n expectations of agents over x, $G_n x=\int_0^1 (E_{in}x)\, di$. Integrating (2.1) over i implies that

$$y_n=G_n\alpha-\beta G_n y_n. \tag{2.1$'$}$$

When $\alpha=0$ and $\beta=-1$ in (2.1), $y_{in}=E_{in}y_n$ so we have the [18] beauty contest problem where each agent seeks to forecast the average action of all the agents. Note that the Nash equilibrium for this model is the action $y_{in}=\alpha/(1+\beta)$ for all i.

At the end of each period n agents observe the price level p_n and update their hierarchies of beliefs over θ. At date $n+1$ they again choose actions via (2.1) but with their revised hierarchies of beliefs. Suppose that the above decision-problem occurs at each date $n=1, 2, \ldots$. Take the error process $\{\varepsilon_n\}_{n=1}^{\infty}$ to be independent and identically distributed with zero mean and bounded first moment.

We shall suppose that the agents do *not* observe the aggregate output at the end of the period. Hence, the learning problem facing agents is not a standard statistical inference problem of determining the coefficients of a linear regression equation from observations of both the regressors and the dependent variable observable.

We shall show that for this model when $|\beta| < 1$, then along each sample path the optimal actions of each agent, y_{in}, will converge to the Nash equilibrium output level, $\alpha/(1+\beta)$, for the true parameter value α. This section will illustrate the main ideas only. The rest of the paper will present the formalism and show how the result in this example is generalized to non-linear multi-dimensional models. The convergence result for this example will have three steps. The formal part of this paper will mimic these three steps.

Step I (*Hierarchies of beliefs over fundamentals*): Recall that for a random variable x, $G_n x$ is the "average opinion" of x, $G_n x \equiv \int_0^1 (E_{in} x)\, di$. Since agent i will not in general know the beliefs of agents $j \neq i$, agent i will not know the value of $G_n x$ so will form an expectation, $E_{in} G_n x$ of the unknown quantity $G_n x$. Define $G_n(G_n x)$ or $G_n^2 x$ to be the average opinion $G_n x$; i.e., $G_n^2 x = \int_0^1 (E_{in} G_n x)\, di$. Inductively, define $G_n^r x$ to be the r-times average opinion of the average opinion ... of x.

There is 1-level knowledge of (MB) if agents know that other agents engage in (MB); there is 2-level knowledge of (MB) if agents know that other agents know that other agents engage in (MB); etc. It is easy to see (by repeatedly substituting (2.1′) into (2.1)) that the following are the implications of such levels of knowledge of (MB):

(1-level knowledge of (MB))

$$y_{in} = E_{in}\alpha - \beta E_{in} G_n \alpha + \beta^2 E_{in} G_n y_n; \tag{2.2}$$

(2-level knowledge of (MB))

$$y_{in} = E_{in}\alpha - \beta E_{in} G_n \alpha + \beta^2 E_{in} G_n^2 \alpha - \beta^3 E_{in} G_n^2 y_n; \tag{2.3}$$

and

(R-level knowledge of (MB))

$$y_{in} = \sum_{r=1}^{R+1} (-\beta)^{r-1} E_{in}(G_n^{r-1}\alpha) + (-\beta)^{R+1} E_{in} G_n^R y_n. \tag{2.4}$$

In (2.4) $G^0\alpha \equiv \alpha$. For example under 2-levels of knowledge of (MB) the optimal action of agent i can be written in terms of that agent's expectation of α, that agent's expectation of the average opinion of α, $E_{in} G_n \alpha$, and that agent's expectation of the average opinion of the average opinion of y_n, $E_{in} G_{in}^2 y_n$.

Now suppose that (i) $0 < \beta < 1$ and (ii) there are some constants $\bar{\alpha}$ and K in $[0, \infty)$, such that it is common knowledge that (α, y_n) lies in

$[0, \bar{\alpha}] \times [-K, K]$. I.e., each agent assigns probability one to the event that (α, y_n) lies in $[0, \bar{\alpha}] \times [-K, K]$; each agent assigns probability one to other agents having beliefs of this form; each agent assigns probability one to other agents having beliefs about other agents belief of the just mentioned form; etc. Take limits as $R \to \infty$ in (2.4) to obtain:

(Common knowledge of (MB))

$$y_{in} = \sum_{r=1}^{\infty} (-\beta)^{r-1} E_{in}(G_n^{r-1}\alpha) \qquad (\text{where } G^0\alpha = \alpha). \tag{2.5}$$

If in addition to (2.5) above the parameter vector α is "common knowledge" (or alternatively, there is only one possible value of the vector α) then (2.5) reduces to $y_{in} = \alpha/(1+\beta)$. This is the Nash equilibrium or rational expectations equilibrium for this model. In particular, under common knowledge of (MB) *and* α, agents will choose the Nash equilibrium output at each date n. This conclusion, for the case when α is common knowledge, is of course the usual Rationalizability argument as in [3], [29], and [13]. When the parameter α is not common knowledge, then the analogous rationalizability argument results in (2.5).

Step II (*Convergence*): Let $(\Omega, \mathcal{F})$ denote the underlying probability space. Let μ_i denote the ex ante belief of agent i over Ω. Assume that for all i and j in I, μ_i is mutually absolutely continuous with respect to each μ_j (i.e., they assign probability zero to the same sets). Let $\mathcal{F}_{i0}$ be the σ-algebra representing any initial private information that agent i may have. Let $\mathcal{F}_{in}$, $n = 1, 2, ...,$ be the σ-algebra representing the information of agent i up to date n–that is, from the observations in $\mathcal{F}_{i0}$ and from the data $\{p_1, ..., p_{n-1}\}$. Agent i's belief at the beginning of date n will therefore be represented by the conditional probability $\mu_i(\cdot \mid \mathcal{F}_{in})$. The expectations operator of agent i, E_{in}, is defined via the probability $\mu_i(\cdot \mid \mathcal{F}_{in})$.

Since $E_{in}[E_{in+1}\alpha] = E_{in+1}\alpha$ the sequence of expectations, $\{E[\alpha \mid \mathcal{F}_{in}]\}_{n=1}^{\infty}$, is a martingale sequence. The martingale convergence theorem then implies that this sequence converges with probability one. This in turn implies that $G_n\alpha \equiv \int_0^1 (E_{in}\alpha)\, di$ converges, and hence that $E_{in}(G_n\alpha)$ converges. Inductively all terms of the form $E_{in}(G_n^{r-1}\alpha)$ will converge. Since $|\beta| < 1$, this in turn implies that the sum in (2.5) converges as $n \to \infty$, hence we may identify on each sample path a limiting value of the actions of each agent, $y_{i\infty} \equiv \lim_{n\to\infty} y_{in}$. (The details of these arguments of course appear in the formal part of the paper.)

Step III (*Identifying the limit*): Fix a sample path. We may write the demand equation as $\bar{p}_n = \alpha - \beta\bar{y}_n + \bar{\varepsilon}_n$, where a bar, "$-$", on top of a variable means the time average (e.g., $\bar{p}_n = \sum_{m=1}^{n} p_m/n$). From the strong law of large numbers we may set $\lim_{n\to\infty}\bar{\varepsilon}_n=0$. Then $\bar{p}_\infty=\alpha-\beta y_\infty$. This in turn implies that

$$y_{i\infty}=\lim_{n\to\infty} y_{in}=\lim_{n\to\infty} E_{in}[\alpha-\beta y_n]=E_{i\infty}[a-by_\infty]=E_{i\infty}\bar{p}_\infty=\bar{p}_\infty, \quad (2.6)$$

where $E_{i\infty}$ is the expectations operator under the limiting information $\bigvee_{n=1}^{\infty}\mathscr{F}_{in}$, and where the last equality follows from the fact that agents will know $\bar{p}_\infty$ in the limit. Upon integrating (2.6) with respect to i we may conclude that $y_\infty=\bar{p}_\infty$. Combining this with the previous conclusion that $\bar{p}_\infty=\alpha-\beta y_\infty$, implies that $y_\infty=\bar{p}_\infty=\alpha/(1+\beta)$. So from (2.6) we conclude that in the limit each agent i chooses the action $y_{i\infty}=\alpha/(1+\beta)$, the Nash equilibrium value for the true value of the parameter α.

2.1. *Summary of the Paper.*

The above concludes the discussion of the example. We now proceed to the formal part of the paper. In Section 4 we introduce the concept of a Savage–Bayesian type. This notion of a type specifies an agent's utility parameters as well as that agent's belief hierarchy over *both* the attribute vectors *and* the actions of other agents at each date. In Section 5 we introduce a contraction property on the optimal decision rules of agents. This plays the role played by the assumption that $|\beta|<1$ in the example of this section. We then show that under the contraction property each agent's optimal action is a function of that agent's belief hierarchy over *only* the attribute vector (and we refer to this belief hierarchy as a Harsanyi type). This step is equivalent to STEP I in the example. In Section 7 we present formally the martingale argument which proves that the belief hierarchies over the attribute vector, and hence the optimal actions, converge over time. This is STEP II of the example. Finally in Section 8 we identify the limiting actions as Nash equilibria to the true game (characterized by the true attribute vectors). This corresponds to STEP III in the example of this section.

In Section 9 we indicate what goes wrong when our key assumptions fail. The failure of the contraction property may lead to an "anything is possible" —any time path of actions is possible under optimizing behavior of agents. We show by example that when the mutual absolute continuity condition on beliefs fails, we may get cyclical behavior on each sample path ad infinitum. All proofs are relegated to the appendix.

3. SOME TERMINOLOGY AND MATHEMATICAL PRELIMINARIES

I is the *finite* set of economic agents. Nature is agent 0, and is not a member of I. Given any collection of sets $\{X_i\}_{i\in I}$, we define $X\equiv\prod_{i\in I}X_i$

and $X_{-i} \equiv \prod_{j \neq i} X_j$ unless otherwise stated; given X_0 and $\{X_i\}_{i \in I}$, we shall sometimes state that $X_{-i} \equiv X_0 \times \prod_{j \neq i} X_j$. Given any collection of functions $f_i: X_i \to Y_i$ for $i \in I$, $f_{-i}: X_{-i} \to Y_{-i}$ is defined by $f_{-i}(x_{-i}) \equiv \prod_{j \neq i} f_j(x_j)$. The cartesian product of metric spaces will always be endowed with the product topology. Let X be any metric space. $\mathscr{P}(X)$ denotes the set of probability measures on X (with X endowed with its Borel σ-algebra, generated by the open sets of X). The set $\mathscr{P}(X)$ will be endowed with the weak topology of measures; (see [4] for more on this). For ease of exposition, wherever the intent is obvious we shall assume, *without mentioning this*, that generic sets and functions are Borel-measurable and generic conditional probabilities are fixed regular versions.

4. THE FORMAL MODEL AT SOME FIXED DATE N

4.1. *Actions and Attribute Vectors*

There is a continuum of agents indexed by the unit interval $I = [0, 1]$, and uniformly distributed on that interval. For technical reasons we shall suppose that these agents may be divided into finitely many classes with each agent in a given class identical in all respects. Each agent i is characterized by an attribute vector θ_i in some space Θ_i, which specifies those parts of agent i's utility function that may be unknown to other agents in the economy. Nature has an attribute vector θ_0 in some space Θ_0, which determines the underlying exogenous randomness of the economy. Θ_0 and Θ_i are assumed to be compact metric spaces. Fix any date n. Agent i must choose at date n an action a_{in} in an *action space* A_i, assumed to be a compact subset of $\mathbb{R}^K$ for some $K < \infty$. The utility function of agent i is the continuous and uniformly bounded function $u_i: \Theta_i \times \Theta_0 \times A_i \times A_{-i} \to \mathbb{R}$, which yields agent i's utility as a function of her own attribute vector $\theta_i \in \Theta_i$, nature's attribute vector $\theta_0 \in \Theta_0$, her own action, $a_i \in A_i$ and the actions of the other agents $a_{-i} \in A_{-i}$. Note that the parameters θ_i and θ_0 are both assumed to be independent of the date.

4.2. *Beliefs and Savage–Bayesian Types*

At the beginning of date n agent i will be characterized by a *Savage–Bayesian type* $q_{in} \equiv (\theta_i, q_{in}^{\infty})$. Here θ_i is agent i's attribute vector; and $q_{in}^{\infty} = (q_{in}^1, q_{in}^2, \ldots)$ is agent i's hierarchy of beliefs over the space of attribute vectors *and* date n action of the agents, $\Theta \times A$. In particular, q_{in}^1 is agent i's first order belief and is a probability which specifies agent i's beliefs over $\Theta_{-i} \times A_{-i}$; q_{in}^2 is agent i's second order belief which specifies agent i's belief about the first order beliefs of other agents; more generally, q_{in}^{r+1} is agent i's $(r+1)$th order belief, which specifies agent i's belief about the rth order beliefs of other agents. Formally, q_{in}^{∞} is an element of a space Q_{in}^{∞} which is as constructed in Appendix A by setting $\Xi_i = \Theta_i \times A_i$ and $\Xi_0 \equiv \Theta_0$ in that construction. $Q_{in} \equiv \Theta_i \times Q_{in}^{\infty}$ is the *space* of agent i's date n Savage–Bayesian types. From Appendix A we know that each hierarchy of beliefs $q_{in}^{\infty} \in Q_{in}^{\infty}$ of agent i induces a unique probability $P_i(q_{in}^{\infty})$ over the space $\Theta_0 \times Q_{-in}$ representing agent i's belief about nature's attribute and the Savage–Bayesian types of other agents.

One may ask: In specifying an agent's beliefs why do we need to go beyond the first level belief, q_{in}^1 over $\Theta_{-1} \times A_{-i}$? After all, is this not all that we require for determining agent i's optimal action? Well, to make statements about "common knowledge of maximizing behavior" we need higher order beliefs. As indicated in the example of Section 2, such common knowledge assumptions will be critical in obtaining our convergence result.

4.3. *Decision Rules*

An agent i with Savage–Bayesian type $q_{in} = (\theta_i, q_{in}^{\infty})$ that chooses an action a_i will receive a date n expected utility equal to

$$\int [u_i(\theta_i, \theta_0, a_i, a_{-i})]\, dP_i(q_{in}^{\infty}), \tag{4.1}$$

where the integration is over $\theta_0 \in \Theta_0$ and $a_{-i} \in A_{-i}$ with respect to the measure $P_i(q_{in}^{\infty})$. Of course, for any $q_{in}^{\infty} = (q_{in}^1, q_{in}^2, \ldots)$, integration over $\Theta_0 \times A_{-i}$ with respect to $P_i(q_{in}^{\infty})$ is the same as integration with respect to the first order belief q_{in}^1.) A *date n decision rule* for agent i is a (Borel-measurable) function $f_{in}\colon Q_{in} \to A_i$ which specifies a rule for how agent i chooses a date n action as a function that her date n Savage–Bayesian type, q_{in}. Define F_{in} to be the set of all such date n decision rules for agent i. That is,

$$F_{in} \equiv \{ f_{in}\colon Q_{in} \to A_i \text{ such that } f_{in} \text{ is Borel-measurable}\}. \tag{4.2}$$

Define $F_{-in} \equiv \prod_{j \neq i} F_j$, the set of tuples of decision rules for all agents other than i, and define $F_n \equiv \prod_{j \in I} F_j$, the set of tuples of decision rules for all agents. We define the set of actions which maximize the expected utility of the agent i with Savage–Bayesian type $q_{in} = (\theta_i, q_{in}^{\infty})$ to be

$$f_{in}^*(q_{in}) \equiv \operatorname*{Argmax}_{a_i \in A_i} \int [u_i(\theta_0, \theta_i, a_i, a_{-i})]\, dP_i(q_{in}^{\infty}). \tag{4.3}$$

To ensure that f_{in}^* is well-defined we impose the following assumption:

$$\begin{array}{l}\text{there is a unique solution to the maximization problem in (4.3)}\\ \text{so that } f_{in}^* \text{ is unique-valued; further, the mapping } f_{in}^*\colon Q_{in} \to A_i\\ \text{defined by (4.3) is Borel-measurable.}\end{array} \tag{4.4}$$

Under the assumption of compactness of A_i and continuity of u_i on A_i, a solution to the maximization problem in (4.3) exists. If we suppose in addition that u_i is strictly concave in a_i, then f^*_{in} will be unique-valued and (from Berge's maximum theorem) will be continuous, and hence Borel-measurable, on Q_{in}. Assumption (4.4) then holds.

Suppose that agent i believes that agents $j \neq i$ are choosing actions via some decision rules specified by $f_{-in} = \{f_{jn}\}_{j \neq i} \in F_{-in}$. If agent i is of Savage–Bayesian type $q_{in} = (\theta_i, q^\infty_{in})$, then her best-response may be denoted by

$$\Psi_{in}(f_{-in})(q_{in}) \equiv \operatorname*{Argmax}_{a_i \in A_i} \int u_i(\theta_0, \theta_i, a_i, f_{-in}(q_{-in}))\, dP_i(q^\infty_{in}), \qquad (4.5)$$

where the integration is over θ_0 and q_{-in} with respect to the measure $P_i(q^\infty_{in})$ induced by agent i's Savage–Bayesian type. Equation (4.5) of course defines a mapping $\Psi_{in}: F_{-in} \to F_{in}$ with the interpretation that if at date n agent i of Savage–Bayesian type q_{in} believes that other agents are using decision rules $f_{-in} \in F_{-in}$ then the action $a_{in} = \Psi_{in}(f_{-in})(q_{in})$ is optimal in the sense implicit in (4.5). Define $\Psi_n: F_n \to F_n$ by setting $\Psi_n(f_n) \equiv \prod_{i \in I} \Psi_{in}(f_{-in})$ for all $f_n \equiv \{f_{in}\}_{i \in I} \in F_n$. $\Psi_n(f_n)$ is the *profile* of best-response decision rules, one for each agent, that the agents use when they believe that others are using the decision rules specified by $f_n = \{f_{in}\}_{i \in I}$.

Notice that the definition of Ψ_{in} above involves a counter-factual of some sort. Agent i of Savage–Bayesian type q_{in} has a belief about the *actions* of other agents given by the measure $P_i(q^\infty_{in})$. In *computing* $\Psi_{in}(f_{-in})$ however we perform the thought-experiment "*if* agent i believes others are using decision rules f_{-in}, what is agent i's optimal action." In reality of course agent i need not believe that others are using those particular decision rules.

In reality, each agent i believes agent j chooses actions via her *optimal* decision rule, f^*_{jn}. We may assume that each agent i knows f^*_{jn} for the following reason: By assumption the parameter θ_j represents that part of agent j's utility function that other agents may have imperfect information about; hence agent i knows j's utility function as a function of θ_j. Further, agent j's belief is uniquely specified by q^∞_{in}. Hence, each agent i can determine the *optimal* action of any agent j *as a function of* that agent j's Savage–Bayesian type $q_{jn} = (\theta_j, q^\infty_{jn})$. We suppose agents do indeed choose optimal actions given their belief and this fact is common knowledge. This implies that each agent will know how other agents are choosing their actions *as a function of* their Savage–Bayesian types. In particular each agent i will know that agent j chooses her action via the decision rule f^*_{jn} of (4.3). Since each agent i knows that the other agents are using decision rules $f^*_{-in} \equiv \{f^*_{jn}\}_{j \neq i}$, and since each such agent i is maximizing expected utility, f^*_{in} must be a best-response to f^*_{-in}, so $f^*_{in} = \Psi_{in}(f^*_{-in})$. This in turn implies that $f^*_n \equiv \{f^*_{in}\}$ is a *fixed point* of the mapping Ψ_n.

5. STEP I: DECISION RULES AS FUNCTIONS OF BELIEF HIERARCHIES OVER FUNDAMENTALS

5.1. We endow the space of agent i's decision rules, F_{in}, with the sup norm, $\|\cdot\|$, which is defined for any $f_{in} \in F_{in}$ by

$$\|f_{in}\| = \text{Sup}\{|f_{in}(q_{in})| \text{ such that } q_{in} \in Q_{in}\}, \tag{5.1}$$

and where $|\cdot|$ is the Euclidean norm over $A_i \subseteq \mathbb{R}^K$. The sup norms on the product spaces F_{-in} and F_n are defined to be the "maximum" of the sup norms of the coordinate spaces; i.e., given any $f_{-in} \in F_{-in}$ (resp. $f_n \in F_n$) define $\|f_{-in}\| \equiv \text{Max}_{j \neq i} \|f_{jn}\|$ (resp. $\|f_n\| \equiv \text{Max}_{i \in I} \|f_{in}\|$). We will now require the best response mapping Ψ_{in} to be a contraction:

(Contraction) For each $i \in I$, $\Psi_{in}: F_{-in} \to F_{in}$ is a contraction of modulus γ_i where $0 \leqslant \gamma_i < 1$. That is, for all f_{-in} and f'_{-in} in F_{-in}, $\|\Psi_{in}(f_{-in}) - \Psi_{in}(f'_{-in})\| \leqslant \gamma_i \|f_{-in} - f'_{-in}\|$. (5.2)

Under (5.2) it is easy to verify that $\Psi_n \equiv \{\Psi_{in}\}_{i \in I}: F_n \to F_n$ is also a contraction operator, of modulus $\gamma = \text{Max}_{i \in I} \gamma_i < 1$. Since F_n is easily verified to be a complete normed space under the sup norm we conclude from the contraction mapping theorem that Ψ_n has a unique fixed point. We argued earlier that the collection of optimal decision rules, $f^*_n \equiv \{f^*_{in}\}_{i \in I}$ is a fixed point of Ψ_n. Hence under (5.2), $f^*_n \equiv \{f^*_{in}\}_{i \in I}$ is the *only* fixed point of Ψ_n.

5.2. *Example* (*The Linear Muth Model*)

We now verify that in the example of Section 2 when the intercept of the demand curve, β, is less than one in absolute value then the contraction property (5.2) holds. In that example we set the optimal action of agent i to be $y_{in} = E_{in}[\alpha - \beta y_n]$ where α is the parameter vector and $y_n = \int_0^1 y_{jn}\, dj$ is the aggregate output level. In verifying that (5.2) we will make life *harder* for ourselves by complicating the model in two ways. First suppose (as we should have in that example) that the optimal action of each agent is $y_{in} = \text{Max}\{0, E_{in}[\alpha - \beta y_n]\}$, so that the agent never chooses negative output and never has negative profit. Suppose also that the parameter β may be *unknown.* The earlier assumption that $|\beta| < 1$ now becomes: there exists a constant $\bar{\beta} < 1$ such that it is common knowledge that the absolute value of β is less than $\bar{\beta} < 1$.

The set up of Section 4 requires that the action spaces be compact. To obtain this we proceed as follows: We suppose that the parameter α is uniformly bounded and in particular that there exists an $\bar{\alpha} > 0$ such that it is common knowledge to all agents that $\alpha \leqslant \bar{\alpha}$. By assumption agents are choosing non-negative outputs. If this fact is known to each agent then the optimal action of each agent becomes $y_{in} = E_{in}[\alpha - \beta y_{in}] \leqslant \bar{\alpha}$. Hence we may take the action space of each agent to be the interval $[0, \bar{\alpha}]$. (The requirement that the optimal actions lie in a compact set is referred to by [13] as the "credible price boundedness" assumption.)

We now verify that the contraction property (5.2) holds. Suppose that agent i believes that each agent $j \neq i$ is using the decision rule $Y_j\colon Q_{jn} \to \mathbb{R}^1_+$; i.e., agent i believes that the agent j of Savage–Bayesian type q_{jn} chooses action $Y_j(q_{jn})$. Then the optimal action for agent i of Savage–Bayesian type q_{in} is the action

$$Y_i(q_{in}) \equiv \text{Max}\left\{0, \int [\alpha - \beta \bar{Y}_{-i}(q_{-in})]\, dP_i(q_{in})\right\}, \tag{5.3}$$

where $\bar{Y}_{-i}(q_{-i}) \equiv \int_0^1 Y_j(q_{jn})\, dj$ is the aggregate output (where, recall, agent i is assumed infinitesmal) and where the integration in (5.3) is over α, β and q_{-in} with respect to the measure $P_i(q_{in})$ induced by agent i's Savage–Bayesian type. Let $\{Y'_j\}_{j \neq i}$ be another collection of decision rules for agents other than i and let $Y'_i(q_{in})$ be the best-response analogous to (5.3). Then, noting that for any two real numbers m and m', $|\text{Max}\{0, m\} - \text{Max}\{0, m'\}| \leqslant |m - m'|$, the following computation shows that the contraction property (5.2) above holds with modulus of contraction $\bar{\beta}$:

$$\|Y_i - Y'_i\| \equiv \sup_{\{q_{in} \in Q_{in}\}} |Y_i(q_{in}) - Y'_i(q_{in})|$$

so from (5.3) and the definition of Y_{-i},

$$\leqslant \bar{\beta} \sup_{q_{in} \in Q_{in}} \int \left[\int_0^1 |Y_j(q_{jn}) - Y'_j(q_{jn})|\, dj\right] dP_i(q_{in})$$

$$\leqslant \bar{\beta} \max_{j \neq i,\, j \in I} \sup_{\{q_{jn} \in Q_{jn}\}} |Y_j(q_{jn}) - Y'_j(q_{jn})|$$

$$= \bar{\beta} \max_{j \neq i,\, j \in I} \|Y_j - Y'_j\| = \bar{\beta}\, \|Y_{-i} - Y'_{-i}\|.$$

5.3. *Example* (*The Non-Linear Muth Model*)

We now consider a non-linear version of the above Muth problem. To this effect let us suppose that the cost function is not quadratic and the demand function is not linear. The objective of the firm may be taken to be to maximize expected profit,

$$E_{in}[p_n y_{in} - c(y_{in})] \text{ s.t. } y_{in} \geqslant 0. \tag{5.4}$$

The first order condition for this problem will be $E_{in}p_n = \partial c(y_{in})/\partial y_{in}$. Hence in general the solution will be of the form $y_{in} = \text{Max}\{0, L(E_{in}p_n)\}$ where L is some possibly non-linear function of the expected price. Next, suppose that the demand curve is of the form $p_n = W(\alpha - \beta y_n) + \varepsilon_n$ where W is a possibly non-linear function of the expected output $\alpha - \beta y_n$ given the parameter vector $\theta_0 = (\alpha, \beta)$ and the aggregate output $y_n \equiv \int y_{jn}\, dj$; and where ε_n is a zero mean shock term. Combining these relations implies that

$$y_{in} = \text{Max}\left\{0, L\left(E_{in}\left[W\left(\alpha - \beta \int_0^1 y_{jn}\, dj\right)\right]\right)\right\}. \tag{5.5}$$

Let $\|L\|$ and $\|W\|$ be the modulus of contraction of the functions L and W, resp. Let $\bar{\beta}$ be a bound on the *unknown* parameter β as in Example 5.2. Assume that $\bar{\beta}\, \|L\| \cdot \|W\| < 1$ (and note that for this L and W need *not* necessarily be contraction operators of modulus less than one). Let $Y_i(q_{in})$ and $Y_{-i}(q_{-in})$ have the same interpretation as in Example 5.2. Then, following the arguments used in Example 5.2, it is easy to verify that $\|Y_i - Y_i'\| \leqslant \bar{\beta}\, \|L\| \cdot \|W\|\, \|Y_{-i} - Y'_{-i}\|$, so that the contraction property holds.

5.4. *Harsanyi Types*

Agent i of Savage–Bayesian type q_{in} will also have a date n belief hierarchy $\tau_{in}^{\infty} = (\tau_{in}^1, \tau_{in}^2, \ldots)$ over the attribute vector $\theta = (\theta_0, \{\theta_i\}_{i \in I})$. In particular, τ_{in}^1 is agent i's first order belief over Θ and is a probability that specifies agent i's beliefs over $\Theta_{-i} \equiv \Theta_0 \times \prod_{j \neq i} \Theta_j$; τ_{in}^2 is agent i's second order belief over Θ which specifies agent i's belief about the first order beliefs of other agents; more generally, τ_{in}^{r+1} is agent i's $(r+1)$th order belief over Θ, which specifies agent i's belief about the r-th order beliefs over Θ of the other agents. Formally, $\tau_{in}^{\infty} = (\tau_{in}^1, \tau_{in}^2, \ldots)$ is an element of a space T_{in}^{∞} as constructed as in Appendix A with $\Xi_0 = \Theta_0$ and $\Xi_i = \Theta_i$ for all $i \in I$. The tuple $\tau_{in} \equiv (\theta_i, \tau_{in}^{\infty})$ consisting of agent i's attribute vector and belief hierarchy over θ is agent i's date n *Harsanyi type*, and is an element of the space of Harsanyi types, $T_i \equiv \Theta_i \times T_i^{\infty}$. Each Harsanyi type $\tau_{in} = (\theta_i, \tau_{in}^{\infty})$ has associated with it a unique probability $p_i(\tau_{in}^{\infty})$ over $\Theta_0 \times T_{-in}$ representing the belief of agent i of Harsanyi type τ_{in}^{∞} about the attribute of nature and the Harsanyi types of other agents.

Agent i's *Savage–Bayesian* type specifies a belief hierarchy over *both* the actions a_{-in} *and* the attribute vectors, θ_{-i}, of other agents. Agent i's

Harsanyi type specifies a belief hierarchy over *only* the attribute vector θ. Via a kind of "projection" operation, each Savage–Bayesian type q_{in} induces a unique Harsanyi type representing tuple of the attribute vector and belief hierarchy over Θ of the agent i with Savage–Bayesian type q_{in}. We denote this $h_i(q_{in})$. For example, the first order coordinate of agent i's Harsanyi type, i.e., τ^1_{in}, is that agent's belief over Θ_{-i}, which is of course the same as the marginal over Θ_{-i} of the first order coordinate of that agent's Savage–Bayesian type, q^1_{in}. (See Appendix B for details. Also note that optimality is not used at all at this stage—only a projection operation.)

Define $C_i(T_{in})$ to be the set of all *continuous* Harsanyi-type based decision rules for agent i:

$$C_i(T_{in}) \equiv \{g_{in}: T_{in} \to A_i \text{ and } g_{in} \text{ is continuous}\}. \tag{5.6}$$

Also define $C_{-i}(T_{-in}) \equiv \prod_{j \neq i} C_j(T_{jn})$ and $C(T_n) \equiv \prod_{i \in I} C_{in}(T_{in})$ where $T_n \equiv \prod_{i \equiv I} T_{in}$. In comparison, recall that F_{in} is the set of decision rules of agent i which are (measurable) functions of agent i's Savage–Bayesian type. Each $g_{in} \in C_i(T_{in})$ may be identified with a unique element f_{in} of F_{in} by setting for each $q_{in} \in Q_{in}$, $f_{in}(q_{in}) = g_{in}(\tau_{in}(q_{in}))$, where $\tau_{in}(q_{in})$ is the Harsanyi type of an agent with Savage–Bayesian type q_{in}. With this identification we shall consider $C_i(T_{in})$ to be a subset of F_{in}.

Suppose agent i believes that all agents $j \neq i$ are choosing actions via Harsanyi-type based decision rules $g_{jn} \in C_j(T_{jn})$. The beliefs of agent i of Harsanyi type $\tau_{in} = (\theta_i, \tau^\infty_{in})$ about the Harsanyi types of other agents is given by probability $p_i(\tau^\infty_{in})$ defined earlier. So if that agent i chooses action a_i her expected utility is

$$\int u_i(\theta_0, \theta_i, a_i, g_{-in}(\tau_{-in}))\, dp_i(\tau^\infty_{in}), \tag{5.7}$$

where the integration is over $\theta_0 \in \Theta_0$ and $\tau_{-i} \in T_{-i}$ with respect to the measure $p_i(\tau^\infty_{in})$. Notice that this expected utility is a function only of agent i's Harsanyi type τ_{in}. Hence the optimal action a_i *may* be chosen as a function of τ_{in}. Under (4.4) the optimal action of an agent from any belief is unique, so the optimal action of each agent *will* be chosen as a function of τ_{in}. From the assumption that the utility function is continuous we may conclude that the resulting decision rule is a *continuous* function of τ_{in}. (See the maximum theorem of [2].) In particular,

$$\Psi_{in}: C_{-i}(T_{-in}) \to C_i(T_{in}) \qquad \text{for all} \quad i \in I. \tag{5.8}$$

This in turn implies that $\Psi_n \equiv \{\Psi_{in}\}_{i \in I}: C(T) \to C(T)$. We argued earlier that Ψ_n is a contraction operator with fixed point equal to the profile of optimal decision rules $f_n^* \equiv \{f_{in}^*\}_{i \in I} \in F_n$. Since $C(T_n)$ is a closed normed space when endowed with the sup norm, the fixed point of Ψ_n lies in $C(T_n)$. Hence we have:

PROPOSITION 5.1. *For each $i \in I$, the optimal decision rule of each agent is a continuous function of that agent's Harsanyi type*; *i.e.* $f_{in}^* \in C_i(T_{in})$.

5.5. *Example* (*The Muth Model*):

Consider again the example of Section 2. There, the attribute vector is the parameter α. An agent's Harsanyi type specifies that agent's belief hierarchy over α. We showed in Eq. (2.5) that under the common knowledge of maximizing behavior (MB), each agent's action is a function of that agent's hierarchy of beliefs over α. Notice that the expression in (2.5) involves sums of expectations. Since expectations are continuous in the probabilities it is easy to verify that (2.5) is a continuous function of the hierarchy of beliefs over α. In particular, from Eq. (2.5) we see that each agent i's optimal action is an element of $C_i(T_i)$. This verifies Proposition 5.1.

6. THE DYNAMIC MODEL

6.1. Let Ω be the underlying space on which all random variables are defined. (Formally, Ω is the "state space" as constructed in [23], and is a complete and separable metric space). Any $\omega \in \Omega$ specifies the action of each agent at each date $n = 1, 2, \ldots$; it also specifies the attribute vector $\theta = \{\theta_0, \{\theta_i\}_{i \in I}\}$ and any other underlying randomness in the economy. $\mathscr{F}$ denotes the set of Borel subsets of Ω. Each agent i has an ex ante subjective belief, μ_i, which is a probability over $(\Omega, \mathscr{F})$ and represents that agent's prior belief over Ω. For each $i \in I$, there is a non-decreasing sequence of σ-algebras $\{\mathscr{F}_{in}\}_{n=0}^{\infty}$ of $\mathscr{F}$. $\mathscr{F}_{i0}$ represents the information agent i receives at "date 0," information which specifies agent i's attribute vector θ_i. $\mathscr{F}_{in}$ is the σ-algebra representing the information of agent i at the beginning of date n (after date $n-1$ decisions have been made). This information would come from all observations of agent i from periods 0 through $n-1$. At the beginning of date n in the state of the world $\omega \in \Omega$, agent i's belief about the true state of the world is represented by the value at ω of the probability μ_i conditional on $\mathscr{F}_{in}$, $\mu_i(\cdot \mid \mathscr{F}_{in})(\omega)$; her Savage–Bayesian type will be some $q_{in}(\omega) = (\theta_{in}(\omega), q_{in}^{\infty}(\omega)) \in Q_{in}$; her Harsanyi type will be given by $\tau_{in}(\omega) = (\theta_{in}(\omega), \tau_{in}^{\infty}(\omega)) \in T_{in}$; and she will choose the action $a_{in}(\omega) = f_{in}^*(q_{in}(\omega))$.

We shall assume that the collection of subjective ex ante beliefs of agents, $\{\mu_i\}_{i \in I}$ obeys:

Condition (*GH*). μ_i and μ_j are mutually absolutely continuous $\forall i, j \in I$; i.e., $\forall i, j \in I$ and $\forall D \subseteq \Omega$, $\mu_i(D) > 0$ if and only if $\mu_j(D) > 0$.

Condition (GH) requires that agents agree ex ante about the events which have zero probability. Condition (GH) does *not* require the ex post probabilities after receipt of information on their utility attributes, $\mu_i(\cdot \mid \mathscr{F}_{i0})(\omega)$ and $\mu_j(\cdot \mid \mathscr{F}_{i0})$, to be mutually absolutely continuous. It should be clear that if $\mu_i = \mu$ for all i, so that μ is a common prior, then condition (GH) holds. Condition (GH) is therefore weaker than the common prior assumption. We therefore name this "condition (GH)" for "Generalized Harsanyi" common prior condition. To enable us to state our results more succinctly we fix a measure μ^* which is mutually absolute continuous to each of the μ_i. For example we could fix an agent i' and set μ^* equal to $\mu_{i'}$. One may wish to think of the measure μ^* as that of an "outside observer." Alternatively μ^* may be interpreted as the "true distribution" if it is generated by the "true" ex ante distribution of the Savage–Bayesian types of agents and the "true" behavior of the agents.

7. STEP II: CONVERGENCE OF BELIEF HIERARCHIES OVER FUNDAMENTALS

7.1. In Section 5 we concluded that each agent i at each date n chooses a date n action as a continuous function of her Harsanyi type. This corresponds to STEP I in the example of Section 2. A Harsanyi type is a date n belief hierarchy over the attribute vector. Note that the true value of the attribute vector does not vary over time—it's value is fixed at date 0. An agent's Harsanyi type is a belief hierarchy over the attribute vector, and may vary over time because of the information received over time. However in the long run this information has smaller and smaller influence on beliefs about the fixed attribute vector. In this section we shall show that each agent i's date n belief hierarchy over the attribute vector, and in particular that agent's date n Harsanyi type, converges over time (i.e., as $n \to \infty$). Note that this is STEP II of the example of Section 2.

We shall use condition (GH) in obtaining the convergence result. The convergence will hold on a set of sample paths with "probability one". Under condition (GH) the "probability one" is with respect to both μ_i *and* μ^*. The results of this section will *not* specify where beliefs converge to, and in particular the results here will *not* claim that there is convergence to complete information about the true parameter vector; this will be the purpose of the next section, Section 8.

For emphasis, we write $\tau^r_{in}(\omega)$ to denote the date n rth order belief of agent i over the attribute vector *in the sample path* $\omega \in \Omega$. We let $\tau_{i\infty}(\omega)$

denote the hierarchy of beliefs under the limiting information field $\mathscr{F}_{i\infty} = \bigvee_{n=1}^{\infty} \mathscr{F}_{in}$ for agent i at ω. Let wlim denote the operation of taking the limit of a sequence of probability measures in the weak topology of measures. (See [4], for more on this.) Define for each $i \in I$ and each r and $n = 1, ..., \infty$,

$$C_i^r \equiv \{\omega \in \Omega \mid \underset{n \to \infty}{\text{wlim}}\ \tau_{in}^r(\omega) = \tau_{i\infty}^r(\omega)\}, \qquad C^r \equiv \bigcap_{i \in I} C_i^r \text{ and } C \equiv \bigcap_{r=1}^{\infty} C^r. \quad (7.1)$$

The set C_i^r is the set of sample paths where the ith agent's rth order beliefs over θ converge; the set C is the set where all orders of beliefs of each agent converge.

THEOREM 7.1 (Convergence of Beliefs). *Suppose condition* (*GH*) *holds. Then* $\mu^*(C) = 1$.

Theorem 7.1 implies that along each sample path (excluding a set with zero probability) the date n Harsanyi type of each agent, τ_{in}, converges as $n \to \infty$ to some value $\tau_{i\infty}$. From Section 5 we know that the date n action a_{in} is a continuous function of the date n Harsanyi type τ_{in}. Theorem 7.1 therefore implies that a_{in} converges to a limiting action $a_{i\infty}$ along each sample path (excluding a set of sample paths with μ^*, and therefore μ_j for all $j \in I$, probability zero). In particular we have:

COROLLARY 7.2. *Let* $a_{in}(\omega)$ *denote the date* n *action of agent* i *in the sample path or state of the world* ω. *Then* $\mu^*(\{\omega \in \Omega \mid a_{in}(\omega)$ *converges as* $n \to \infty$ *to some* $a_{i\infty}(\omega)\}) = 1$.

8. STEP III: IDENTIFYING LIMIT ACTIONS $\{a_{i\infty}\}_{i \in I}$ AS EQUILIBRIA

8.1. The collection of actions $\{a_i^*\}_{i \in I}$ is said to be a *Nash* (*or Rational Expectations*) equilibrium *for* the attribute vector $\theta = (\theta_0, \{\theta_i\}_{i \in I})$ if $\forall i \in I$,

$$a_i^* \in \underset{\{a_i \in A_i\}}{\text{Argmax}}\ u_i(\theta_0, \theta_i, a_i, a_{-i}^*). \quad (8.1)$$

To characterize the limiting actions as Nash equilibria for the true θ_0, we will have to provide a condition which says that enough information about θ_0 is obtained in the limit. It turns out that we do not need to assume that θ_0 is learnt completely in the limit; a sufficient statistic is enough. Indeed, for each $i \in I$, suppose that there exists a function H_i: $\Theta_0 \times A_{-i} \to \mathbb{R}^m$, for

$m < \infty$ and a function $U_i: \Theta_i \times A_i \times \mathbb{R}^m \to \mathbb{R}$ such that the utility function of agent i may be written as

$$u_i(\theta_i, \theta_0, a_i, a_{-i}) = U_i(\theta_i, a_i, H_i(\theta_0, a_{-i})) \tag{8.2}$$

$\forall(\theta_0, \theta_i, a_i, a_{-i}) \in \Theta_0 \times \Theta_i \times A_i \times A_{-i}$. In (8.2) we see that $H_i(\theta_0, a_{-i})$ is a "sufficient statistic" for θ_0 and a_{-i} when computing agent i's utility. Suppose that each agent i knows the value of $h_i^* = H_i(\theta_0, a_{-i}^*)$. Suppose also that each agent is maximizing utility subject to knowledge of h_i^*; i.e., suppose that

$$a_i^* \in \operatorname{Argmax} U_i(\theta_i, \cdot, h_i^*) \qquad \text{where} \quad h_i^* = H_i(\theta_0, a_{-i}^*). \tag{8.3}$$

Then from the definition of U_i in (8.2), (8.3) implies that (8.1) is true. In particular the actions $\{a_i^*\}_{i \in I}$ are a Nash equilibrium for the true attribute vector $\theta = (\theta_0, \{\theta_i\}_{i \in I})$. Note that an agent may know the true value of $H_i(\theta_0, a_{-i}^*)$ without knowing the individual values of θ_0 and $a_{-i}^* \equiv \{a_j^*\}_{j \neq i}$. In that case the collection of actions $\{a_i^*\}_{i \in I}$ may be a Nash equilibrium in the sense that they are optimal when each agent i behaves *as if* she knows the true θ_0 and the true actions of others, a_{-i}^*; however agent i may in reality choose the action a_i^* based on a belief about θ_0 and the actions of others, a_{-i}^*, which is different from, or involves uncertainty about, the true value of θ_0 and the actions of others.

We know from Section 7 that under our assumptions the actions of agents converge so that $a_{in} \to a_{i\infty}$ for all $i \in I$. If each agent i's utility function has the representation in (8.2) and H_i is continuous then we may conclude that $h_{in} \equiv H_i(\theta_0, a_{-in})$ converges to $h_{i\infty} \equiv H_i(\theta_0, a_{-i\infty})$. We will assume that agent i learns in the limit the value of $h_{i\infty} \equiv H_i(\theta_0, a_{-i\infty})$. The earlier argument then implies that the limiting actions constitute a Nash equilibrium.

The requirement that agent i learns $h_{i\infty}$ in the limit is equivalent to the requirement that $h_{i\infty}$ be measurable with respect to the agent's limit information, $\mathscr{F}_{i\infty} \equiv \bigvee_{n=0}^{\infty} \mathscr{F}_{in}$. We now provide some justification for such an assumption: Note that h_{in} is the item that appears in the agent i's utility function, perhaps after integrating out noise terms. At the end of each period after the utility has been "enjoyed," the agent should be able to invert the utility function and recover a perhaps noisy estimate of h_{in}. Since $h_{in} \equiv H_i(\theta_0, a_{-in})$ and a_{-in} converges, it is reasonable to expect that in the limit the agent will learn the limiting value of h_{in}-usually by averaging out the noise term via a law of large numbers type argument. We show in Section 8.2 that this is the case in the Muth examples. For more general techniques of handling this learning problem, and in particular on determining when h_∞ is $\mathscr{F}_{i\infty}$ measurable, see [26].

The proposition below, our main theorem, summarizes this discussion.

PROPOSITION 8.1. *Suppose the utility function of each agent may be written as* (8.2) *above for some collection* $\{H_i\}_{i \in I}$ *of uniformly bounded and continuous functions. Suppose that for each* $i \in I$, $H_i(\theta_0, a_{-i\infty})$ *is* $\mathscr{F}_{i\infty}$*-measurable* (*or, equivalently, each agent "learns" in the limit the value of* $H_i(\theta_0, a_{-i\infty})$). *Then for* μ^*-a.e. *sample path, the collection of limiting actions,* $\{a_{i\infty}\}_{i \in I}$ *is a Nash* (*or Rational Expectation*) *Equilibrium for the true parameter vector* $\theta = (\theta_0, \{\theta_i\}_{i \in I})$ *defined for that sample path.*

8.2. *Example*

Consider the Muth model of Section 2. In that example nature's attribute vector is the parameter $\theta_0 = \alpha$. We may define

$$H_i(\alpha, y_{-i}) \equiv \alpha - \beta \int_0^1 y_j \, dj, \tag{8.4}$$

which is equal to the expected price when the actions of all agents except i (who is "small") is $y_{-i} = \{y_j\}_{j \neq i}$ and the intercept of the demand curve is α. The profit function of agent i when she chooses output y_i may then be written as $U_i(\theta_i, y_i, H_i(\theta_0, y_{-i})) = H_i(\theta_0, y_{-i}) \, y_i - 0.5 y_i^2$. Hence we obtain the utility function in (8.2) above. In the *nonlinear* Muth model of Section 5.3, a function $W(\cdot)$ determines the expected price so we may set

$$H_i(\theta_0, y_{-i}) \equiv W\left(\alpha - \beta \int_0^1 y_i \, dj\right) \tag{8.5}$$

in which case the profit function becomes $U_i(\theta_i, y_i, H_i(\theta_0, y_{-i})) = \text{Max}\{0, y_i W(\alpha - \beta \int y_j \, d_j) - c(y_i)\}$, and again (8.2) holds.

By assumption each agent observes the price. In the linear Muth model of Section 2 this means that each agent observes $p_n = \alpha - \beta \int_0^1 y_j \, dj + \varepsilon_n$, which from (8.4) becomes $p_n = H_i(\alpha, y_{-in}) + \varepsilon_n$. Taking the average over n of this latter equation and invoking the strong law of large numbers implies that $\bar{p}_\infty = H_i(\alpha, y_{-i\infty})$ where $\bar{p}_\infty = \lim_{N \to \infty} \sum_{n=1}^N p_n$. Hence we see that agent i learns the limiting value, $H_i(\alpha, y_{-i\infty})$, so we may apply Proposition 8.1 in this case. For the non-linear Muth model each agent i observes $p_n = W(\alpha - \beta \int_0^1 y_j \, dj) + \varepsilon_n$. Hence a similar appeal to the strong law of large numbers implies that agent i will learn the limiting value $H_i(\alpha, y_{-i\infty})$, so again Proposition 8.1 may be used to conclude that there is convergence to the Nash equilibrium of the true model.

Consider again the linear Muth model but suppose that there is imperfect information over *both* the parameter α and β as well as the individual actions of the agents $\{y_j\}_{j \in I}$. Let us denote the true values of these by

α^*, β^* and $\{y_j^*\}_{j\in I}$, respectively. Suppose however that each agent knows the true value of quantity $(\alpha - \beta y)$ where $y = \int_0^1 y_j\, dj$, the aggregate output; i.e., suppose that each agent i assigns probability one to the event $\{(\alpha, \beta, \{y_j\}_{j\in I}) \mid \alpha - \beta \int_0^1 y_j\, dj = \alpha^* - \beta^* \int_0^1 y_j^*\, dj\}$. Then, each agent i will choose the action $y_i^* = E_i p = E_i[\alpha - \beta y] = \alpha^* - \beta^* y^*$. Integrating both sides of this equation over i implies that $y^* = \alpha^* - \beta^* y^*$ so $y^* = \alpha^*/(1+\beta^*)$. This in turn implies that $y_i^* = \alpha/(1+\beta)$. Hence each agent is choosing the Nash equilibrium output level. However, this choice of action may be based on imperfect information over α, β, and y! ▮

9. WHAT HAPPENS WHEN KEY ASSUMPTIONS ARE VIOLATED?

9.1. *The Violation of Contraction Property.*

Condition (R) below says that for each action a_i of agent i in some set $\bar{A}_i$, and for each pair (θ_i, τ_i^1) representing agent i's attribute vector, θ_i, and belief about others' attribute vectors, τ_i^1, we can find a belief over actions of others (in $\bar{A}_{-i}$) which "rationalizes" the action a_i given (θ_i, τ_i^1):

Condition (*R*). Fix any subsets $\bar{A}_i \subseteq A_i$ for $i \in I$. Suppose that $\forall i \in I$, and $\forall (\theta_i, \tau_i^1, a_i) \in \Theta_i \times T_i^1 \times \bar{A}_i$, $\exists q_i^1 \in Q_i^1$ such that (i) $a_i = f_i^*(\theta_i, q_i^1)$ and (ii) $h_i^1(q_i^1) = \tau_i^1$.

Consider the linear Muth example with the parameter vector $\theta_0 = (\alpha, \beta)$ is common knowledge and $\beta \geqslant 0$. One can check that there are exactly two situations: (i) $\beta \in [0, 1)$, in which case the contraction property holds and condition (R) can only be satisfied when $\bar{A}_i = \{\alpha/(1+\beta)\}$ for all i, the singleton set containing only the Nash equilibrium action; or (ii) $\beta \geqslant 1$, in which case the contraction property fails but condition (R) holds for $\bar{A}_i = [0, \alpha]$. As a second example, consider the linear Muth example but where there is merely common knowledge that $\theta_0 = (\alpha, \beta)$ lies in some set $[\underline{\alpha}, \bar{\alpha}] \times [\underline{\beta}, \bar{\beta}]$. Suppose that $\underline{\beta} \geqslant \bar{\alpha}/\underline{\alpha}$ (and note that a necessary condition for this is that $\underline{\beta} \geqslant 1$). Define $\bar{A}_i = \bar{A} = [0, \underline{\alpha}]$. One can show that condition (R) holds in this case too. (To see this, fix any $y_i \in \bar{A}_i$, and any (first order) belief of i over (α, β). Suppose further that agent i's expectation of the aggregate output is $E_i y = [E_i \alpha - y_i]/E_i \beta$ (and that i believes y and (α, β) are independent). One can check that $E_i y \geqslant 0$ and $E_i y \leqslant \bar{\alpha}/\underline{\beta} \leqslant \underline{\alpha}$. Hence (R) holds.)

We will show in Theorem 9.1 below that under condition (R) we have an "anything is possible" result: Any stochastic process of actions (in $\prod_{n=1}^{\infty} \prod_{i\in I} \bar{A}_i$) is a possible outcome consistent with common knowledge of maximizing behavior (MB)). The intuition behind Theorem 9.1 is as follows: The only restrictions between date n and date $n+1$ beliefs is due to the fact that the true parameter vector, θ, is fixed between those two periods. This means that we can not vary agents belief hierarchies over θ across time at will. Instead, beliefs will obey a Bayesian updating condition and will converge, from Theorem 7.1. Under the contraction property the belief hierarchy over θ determines completely the optimal actions of agents (Proposition 5.1)—from which we obtain the convergence of actions. Without the contraction property, and in particular with condition (R), there is no such link between belief hierarchies over θ and optimal actions.

To state our anything is possible result we need a bit more structure. Consider as fixed Θ, and $\{A_i, u_i\}_{i \in I}$. For the dynamic problem we need to specify in addition the process governing the observations of agents (which in the case of the Muth example is the price process). To this effect, fix some sets $\{Z_i\}_{i \in I}$ with the interpretation that $z_{in} \in Z_i$ is agent i's end of date n observation. Agent i's end of date n history of observations will therefore be $\{\theta_i, a_{i1}, z_{i1}, ..., a_{in}, z_{in}\}$, which includes her attribute vector θ_i and past history of own actions. We take as given, or as a primitive of the model, the conditional probabilities $P(z_n \mid a_n, z_{n-1}, ..., z_1, a_1, \theta)$ which show how the date n observations are generated as a function of the past actions and observations and the initial date 1 attribute vectors. Define $\wp_{AZ}$ to be the set of all probability distributions over the stochastic process $(\theta, a_1, z_1, a_2, z_2, \ldots)$ such that for all n the conditional probability of z_n given $(a_n, z_{n-1}, ..., z_1, a_1, \theta)$ is $P(z_n \mid a_n, z_{n-1}, ..., z_1, a_1, \theta)$. We then have:

PROPOSITION 9.1 (Anything is Possible). *Suppose that condition* (R) *holds. Fix any stochastic law $P_{AZ} \in \wp_{AZ}$ such that $P_{AZ}(\{\{a_{in}\}_{i \in I} \in \prod_{i \in I} \bar{A}_i$ for all $n\}) = 1$. Then there exist ex ante subjective beliefs of agents $\{\mu_i\}_{i \in I}$ on a state space Ω such that the induced distribution over the process of optimal actions and observations of the agents is precisely that obtained from P_{AZ}.*

Observe that this theorem places no conditions on the true value of the parameter θ_0. For example in the linear Muth model it suffices that agents believe that the true value of β is larger than one for the theorem to hold, even though the true value of β may be strictly less than one.

9.2. *Example Where Agents have Misspecified Priors* (*i.e.*, (*GH*) *Fails*) *and Outputs Cycle on Every Sample Path.*

Consider the linear Muth example parametrized as follows: $\beta = 1/3$ and this fact is common knowledge; and each agent i believes that all other agents j believe that it is common knowledge that $\alpha = 4/3$. Then (2.5) implies that agent i believes that all other agents j will choose action $y_j = \alpha/(1+\beta) = 1$ in each period. Each agent i believes that the parameter α can be one of two points, $\alpha' = 4/3$ or $\alpha'' = 1$. Let v_1 be the probability assigned to the event $\{\alpha = \alpha'\}$ and suppose that $0 < v_1 < 1$. Notice that each agent assigns probability zero to the true belief hierarchies of other agents; each agent i is sure the other agents j will choose actions $y_j = 1$ at all dates while in fact each agent will choose actions which depend upon beliefs of α over $\{\alpha', \alpha''\}$. Indeed, if v_n is agent i's date n probability assigned to the event $\{\alpha = \alpha'\}$ then from (2.1) agent i's optimal date n action is

$$y_{in} = E_{in} p_n = E_{in}[\alpha - \beta y_n] = v_n/3 + 2/3. \tag{9.1}$$

Hence condition (GH) is violated.

Suppose that the true value of α is $\alpha^* = 10/9$. Suppose also that the noise process $\{\varepsilon_n\}_{n=1}^{\infty}$ is normally distributed with zero mean and unit variance. Since ε has support on all of the real line observe that despite the fact that the ith agent's beliefs assign probability zero to the true beliefs of others, the agent is not confronted with data (a price level) at any finite date that can not be explained with some value of the noise term ε at that date. Hence at each date the agent may update her beliefs using Bayes' rule in the usual manner.

For this model the Nash equilibrium output is 5/6. In the proposition below we show that optimal outputs, y_n, cycle ad infinitum between outputs of 1 and 2/3, and hence around the Nash equilibrium output. Also, beliefs v_n cycle between 0 and 1.

PROPOSITION 9.2. *Let μ^* denote the true probability (over Ω) generated by the true value of $\alpha = \alpha^*$. Then for μ^*-a.e. sample path*, $\limsup_{n \to \infty} y_n = 1$, $\liminf_{n \to \infty} y_n = 2/3$, $\limsup_{n \to \infty} v_n = 1$ *and* $\liminf_{n \to \infty} v_n = 0$.

The intuition behind the result is the following: Suppose that over time agent i assigns large probability to the intercept term being equal to α'. Then from (9.1) the agent will choose actions approximately equal to 1. The data observed by the agent is then approximately $p^* = \alpha^* - \beta = 10/9 - 1/3 = 7/9$, ignoring the shock terms. Since 7/9 is closer to $\alpha'' - \beta = 1 - 1/3 = 2/3$ than to $\alpha' - \beta = 4/3 - 1/3 = 1$, the agent will assign more weight to α'' being the true value of α than to α'. As the agent becomes more confident that α'' is the true value of α, from (9.1) the agent will choose actions approximately equal to 2/3. Ignoring the shocks, the data observed by the agent is then approximately $p_n^* = \alpha - (2/3)\beta = 10/9 - 2/9 = 8/9$. The agent believes $y = 1$ so believes that the data generating process is $p_n = \alpha - \beta$. Since 8/9 is closer to $\alpha' - \beta = 4/3 - 1/3 = 1$ than to $\alpha'' - \beta = 1 - 1/3 = 2/3$, the agent will assign more weight to α' being the true value of α. This process then repeats itselfs, and hence we see how optimal actions may cycle between the values $y = 1$ and $y = 2/3$ ad infinitum.

9.3. *The Continuum of Agents and Zero Discount Factor Assumptions*

With our continuum assumption on the set of agents, there is no active learning (or "teaching"). In particular, agents can not individually change their actions to either get better information or to signal information to other agents. The zero discount factor assumption implies that the agents would not even want to do such active learning, even if they could, since all payoffs would acrue in the future and that is not taken into account in current decisions. If one were to introduce positive discount factors and drop the continuum assumption, once could still make some progress using the techniques of [17], [16] and [21]. Several differences have to be noted though. First, the papers just mentioned in general do not obtain convergence to Nash, but rather to subjective equilibrium of the repeated game. Note well that from the folk theorem there are many Nash equilibria to the repeated game—and hence even more subjective equilibria. Second, the above mentioned papers typically require finitely many actions and require agents to observe the same past history of actions—private information is not allowed in those papers. Finally, note that when agents observe the price signal, their "types" become correlated, even if initially they were independent. One would then have to appeal to the results of [22] (which is only for zero discount factors) rather than the above mentioned papers to make progress on the positive discount factor case, and even in that case one may have to settle with correlated equilibria, suitably defined for the positive discount factor model.

10. APPENDIX A: HIERARCHIES OF BELIEFS AND THE SPACE B_i

10.1. In this appendix we sketch the construction of hierarchies of beliefs. The reader should consult [23] for the details and for references to the literature. Recall that I is the set of (*finitely* many classes of) agents and nature is referred to as agent 0 (not a member of I). We take as given a collection of complete and separable metric spaces Ξ_0 and $\{\Xi_i\}_{i\in I}$. We shall consider Ξ_i to be the set pertaining to agent i; this will have the meaning that i "knows" her own value of $\xi_i \in \Xi_i$. We consider Ξ_0 to be the parameters of "nature." We proceed to construct the space of hierarchies of beliefs over the space $\Xi = \Xi_0 \times \prod_{i\in I} \Xi_i$. Recall that given any metric space X, $\mathscr{P}(X)$ denotes the set of probability measures on the Borel subsets of X. Recall also that $\mathscr{P}(X)$ is endowed with the weak topology of measures. If X is a complete and separable metric space then so is $\mathscr{P}(X)$. (See, e.g., [28, Parthasarathy Theorems II.6.2 and II.6.5.])

Construct the sets $\{B_i^r\}_{r=1}^{\infty}$ inductively as

$$B_i^1 \equiv \mathscr{P}(\Xi_{-i}), \qquad \text{where} \quad \Xi_{-i} \equiv \Xi_0 \times \prod_{j \neq i} \Xi_j, \tag{10.1}$$

and given $\{B_j^r\}_{j \in I}$ for some $r \geqslant 1$, define

$$B_i^{r+1} \equiv \mathscr{P}(B_{-i}^r \times \Xi_{-i}). \tag{10.2}$$

An element $b_i^1 \in B_i^1$ represents agent i's belief about $\xi_{-i} \in \Xi_{-i}$ and shall be referred to as agent i's first order belief. An element $b_i^2 \in B_i^2$ specifies agent i's belief about the first order beliefs of others and shall be referred to as agent i's second order belief. An element $b_i^r \in B_i^r$ is i's rth order belief and it specifies agent i's belief about the $(r-1)$th order beliefs of other agents.

It should be clear that higher order beliefs of an agent should be related to the lower order beliefs of the *same* agent by some kind of projection operation. For example, if b_i^1 and b_i^2 are the first and second order beliefs of the same agent then b_i^1 should be the marginal distribution of b_i^2 on Ξ_{-i}. To express this relation we define the functions $\phi_i^r \colon B_i^{r+1} \to B_i^r$ inductively as follows: For any subset $D \subseteq \Xi_{-i}$,

$$\phi_i^1(b_i^2)(D) \equiv b_i^2(\{B_{-i}^1 \times D\}) \qquad \text{for all} \quad b_i^2 \in B_i^2; \tag{10.3}$$

i.e., ϕ_i^1 is the operator that yields the marginal distribution on Ξ_{-i} from any joint distribution on $B_{-i}^1 \times \Xi_{-i}$; and given $\{\phi_j^{r-1}\}_{j \in I}$ define ϕ_i^r by setting for any $b_i^{r+1} \in B_i^{r+1}$ and any $D \subseteq B_{-i}^{r-1} \times \Xi_{-i}$,

$$\phi_i^r(b_i^{r+1})(D) \equiv b_i^{r+1}(\{(b_{-i}^r, \xi_{-i}) \in B_{-i}^r \times \Xi_{-i} \colon (\phi_{-i}^{r-1}(b_{-i}^r), \xi_{-i}) \in D\}). \tag{10.4}$$

The set of all possible belief hierarchies of agent i is then defined to be the set

$$B_i \equiv \left\{ (b_i^1, b_i^2, \ldots) \in \prod_{r=1}^{\infty} B_i^r \colon b_i^r = \phi_i^r(b_i^{r+1}) \text{ for all } r \geqslant 1 \right\}. \tag{10.5}$$

10.2. *The Mapping* $P_i \colon B_i \to \mathscr{P}(B_{-i} \times \Xi_{-i})$

Fix any element $b_i = (b_i^1, b_i^2, \ldots) \in B_i$. We will now construct an "associated" probability $P_i(b_i) \in \mathscr{P}(B_{-i} \times \Xi_{-i})$ with the property that for each integer r, the marginal of $P_i(b_i)$ on $B_{-i}^r \times \Xi_{-i}$ is equal to b_i^{r+1}. Recall that by definition B_{-i} is a subset of the infinite cartesian product $\prod_{r=1}^{\infty} B_{-i}^r$. Define $\text{PROJ}_{-i,r} \colon B_{-i} \times \Xi_{-i} \to B_{-i}^r \times \Xi_{-i}$ to be the projection of the space $B_{-i} \times \Xi_{-i}$ onto $B_{-i}^r \times \Xi_{-i}$. The inverse of the projection mapping, $(\text{PROJ}_{-i,r})^{-1} D^r$, is the "upliftment" of the subset $D^r \subseteq B_{-i}^r \times \Xi_{-i}$ to the infinite cartesian product $B_{-i} \times \Xi_{-i}$. Fix any integer r and define the class Δ^r of r-cylinder subsets of $B_{-i} \times \Xi_{-i}$ to be those which are "upliftments" of some subset of $B^r_{-i} \times \Xi_{-i}$; i.e.,

$$\Delta^r \equiv \{D \subseteq B_{-i} \times \Xi_{-i} : D = (\text{PROJ}_{-i,r})^{-1} D^r \text{ for some (measurable) } D^r \subseteq B^r_{-i} \times \Xi_{-i}\}. \tag{10.6}$$

For any such cylinder set $D \in \Delta^r$ and any $b_i = \{b_i^1, b_i^2, ...\} \in B_i$, define

$$P_i(b_i)(D) = b_i^{r+1}(\{\text{PROJ}_{-i,r} D\}). \tag{10.7}$$

From the probabilistic coherence condition implicit in the definition of B_i in (10.5), it is easy to check that for each $b_i \in B_i$, $P_i(b_i)$ in (10.7) is "well-defined" over $\bigcup_{r=1}^{\infty} \Delta^r$ (in a sense analogous to the Kolmogorov consistency condition), and hence extends to a unique probability measure over all Borel subsets of $B_{-i} \times \Xi_{-i}$. (See for example [28, Theorem 4.2., p. 143] or [23].) Since cylinder sets are "convergence determining" it is easy to check that the mapping $P_i: B_i \to \mathscr{P}(B_{-i} \times \Xi_{-i})$ is *continuous*.

10.3. *The Mapping* $h_i: Q_i \to T_i$

Fix any date n. For expositional convenience we drop the subscript n from Q_{in} and T_{in}. We now define inductively the mapping $h_i: Q_i \to T_i$ which determines the Harsanyi type $\tau_i = h_i(q_i)$ of an agent with Savage–Bayesian type q_i. Define $h_i^1: Q_i^1 \to T_i^1$ by $h_i^1(q_i^1) \equiv \text{Marg}\, q_i^1$, $\forall q_i^1 \in Q_i^1$. Next suppose we have defined for some $r \geqslant 1$ the function $h_i^r: Q_i^r \to T_i^r$.

Define $h_i^{r+1}: Q_i^{r+1} \to T_i^{r+1}$ as follows: For each $q_i^{r+1} \in Q_i^{r+1}$ and for all $S \subseteq \Theta_{-i} \times T^r_{-i}$, $h_i^{r+1}(q_i^{r+1})(S) = q_i^{r+1}(\{(\theta_{-i}, a_{-i}, q^r_{-i}) \in \Theta_{-i} \times A_{-i} \times Q^r_{-i} \mid (\theta_{-i}, h^r_{-i}(q^r_{-i})) \in S\})$. By induction on r we have therefore constructed a sequence of functions $h_i^r: Q_i^r \to T_i^r$ $\forall i \in I$ and $\forall r = 1, 2, ...$ Finally define $h_i: Q_i \to T_i$ by setting for each $q_i = (\theta_i, q_i^1, q_i^2, ...) \in Q_i$, $h_i(q_i) \equiv (\theta_i, h_i^1(q_i^1), h_i^2(q_i^2), ...)$. (It is easy to verify that $h_i(q_i)$ thus defined lies in T_i).

11. APPENDIX B: THE PROOFS

Proof of Theorem 7.1. We begin with the following result, which is a direct implication of the Martingale Convergence theorem (see [10, Theorem 9.4.8., (p. 340]):

Claim. Let X_n for $n = 1, 2, ...$ and $n = \infty$ be a sequence of uniformly bounded real-valued random variables on Ω. Fix any $i \in I$ and suppose that $\lim_{n \to \infty} X_n = X_\infty \mu_i$-a.e. Then $\lim_{n \to \infty} \int X_n \, d\mu_i(\cdot \mid \mathscr{F}_{in}) = \int X_\infty \, d\mu_i(\cdot \mid \mathscr{F}_{i\infty})$, μ_i-a.e.

Using standard arguments it is easy to show that on any complete and separable metric space S, there exists a countable set, $\Gamma(S)$, of uniformly bounded continuous real-valued functions on S which are convergence-determining in the following sense: The weak convergence of a sequence of probability measures $\{m_n\}_{n=1}^{\infty}$ to a measure m_∞ holds whenever $\lim_{n\to\infty} \int g\, dm_n = \int g\, dm_\infty$ for each g in $\Gamma(S)$. Recall the definitions of the sets C_i^k and C^k in (7.1). Fix an $i \in I$ and any real-valued bounded continuous function, g, on Θ. Apply the claim with $X_n = X_\infty = g(\theta)$; then we may conclude that μ_i-a.e., $\lim_{n\to\infty} \int_\Theta g(\theta)\, \mu_i(d\theta \mid \mathscr{F}_{in}) = \int_\Theta g(\theta)\, \mu_i(d\theta \mid \mathscr{F}_{i\infty})$. From the definition of τ_{in}^1 this implies that $\lim_{n\to\infty} \int g\, d\tau_{in}^1 = \int g\, d\tau_{i\infty}^1$, μ_i a.e. This can be made to hold with μ probability one for any countable collection of such functions g (simultaneously) and hence for all $g \in \Gamma(\Theta)$. Hence $w \lim_{n\to\infty} \tau_{in}^1 = \tau_{i\infty}^1$ with μ_i probability one. In particular, $\mu_i(C_i^1) = 1$ for each i. From the definition of μ^* under condition (GH), this implies that $\mu^*(C_i^1) = 1$ for all $i \in I$ and hence $\mu^*(C^1) = \mu^*(\bigcap_{i\in I} C_i^1) = 1$.

Next suppose that for some $r = 1, 2, \ldots$, we have shown that $\mu^*(C^r) = 1$. Fix any real-valued bounded continuous function g on $T^r \times \Theta$. Set $X_n \equiv g(\tau_n^r, \theta)$ for $n = 1, \ldots, \infty$. Then $\mu^*(C^r) = 1$ implies that $\lim_{n\to\infty} X_n = X_\infty$ μ-a.e. From the definition of μ^*, the same is true μ_i-a.e for each $i \in I$. The claim then implies that $\lim_{n\to\infty} \int X_n\, d\mu_i(\cdot \mid \mathscr{F}_{in}) = \int X_\infty\, d\mu_i(\cdot \mid \mathscr{F}_{i\infty})$ μ_i-a.e. From the definition of τ_{in}^{r+1} this implies that $\lim_{n\to\infty} \int g\, d\tau_{in}^{r+1} = \int g\, d\tau_{i\infty}^{r+1}$, μ_i-a.e. This can be made to hold with μ_i probability one for all functions g in the countable set $\Gamma(T^r \times \Theta)$. Hence $w \lim_{n\to\infty} \tau_{in}^{r+1} = \tau_{i\infty}^{r+1}$ μ_i-a.e. So $\mu_i(C_i^{r+1}) = 1$ and therefore $\mu^*(C_i^{r+1}) = 1$. Since this is true for all $i \in I$ we obtain that $\mu^*(C^{r+1}) = \mu^*(\bigcap_{i\in I} C_i^{r+1}) = 1$. By induction we therefore conclude that $\mu^*(C^r) = 1$ for all $r = 1, 2, \ldots$. Hence $\mu^*(C) = 1$. ∎

Proof of Proposition 8.1. Define $h_{in} = H_i(\theta_0, a_{-in})$ and $h_{i\infty} = H_i(\theta_0, a_{-i\infty})$. Since each agent is choosing optimal actions at each date, $\forall \hat{a}_i \in A_i$, $E_i[U_i(\theta_i, a_{in}, h_{in}) \mid \mathscr{F}_{in}] \geqslant E_i[U_i(\theta_i, \hat{a}_i, h_{in}) \mid \mathscr{F}_{in}]$, μ^*-a.e. Taking limits as $n \to \infty$ and applying the claim implies that $\forall \hat{a}_i \in A_i$, $E_i[U_i(\theta_i, a_{i\infty}, h_{i\infty}) \mid \mathscr{F}_{i\infty}] \geqslant E_i[U_i(\theta_i, \hat{a}_i, h_{i\infty}) \mid \mathscr{F}_{i\infty}]$, μ^*-a.e. From the hypothesis of this proposition $h_{i\infty} = H_i(\theta_0, a_{-i\infty})$ is $\mathscr{F}_{i\infty}$-measurable, so $U_i(\theta_i, a_{i\infty}, h_{i\infty}) \geqslant U_i(\theta_i, \hat{a}_i, h_{i\infty})$. But then from the definition of U_i we obtain $u_i(\theta_0, \theta_i, a_{i\infty}, a_{-i\infty}) \geqslant u_i(\theta_0, \theta_i, \hat{a}_i, a_{-i\infty})$, which proves the proposition. ∎

Proof of Proposition 9.1. We begin with the following lemma:

LEMMA 9.1.1. *$\forall i \in I$, $\exists F_i\colon T_i \times \bar{A}_i \to Q_i$ such that if $q_i = F_i(\tau_i, a_i)$ then $a_i = f_i^*(\theta_i, q_i)$ and $h_i(q_i) = \tau_i$.*

Proof of Lemma 9.1.1. First suppose that $\bar{A}_i = A_i$ for each $i \in I$. Define $\forall i \in I$, $D_i^1 \equiv \Theta_i \times T_i^1 \times A_i$, and for $r = 2, 3, \ldots$, $D_i^r \equiv \{(\tau_i^r, q_i^{r-1}) \in T_i^r \times Q_i^{r-1} \mid \phi_i^{r-1}(\tau_i^r) = h_i^{r-1}(q_i^{r-1})\}$. The restriction in the definition above (for $r \geqslant 2$) requires that τ_i^r and q_i^{r-1} have the same associated $(r-1)$th order Harsanyi type. Suppose that for some $r = 1, 2, \ldots$, we have the existence of a function $F_j^r\colon D_j^r \to Q_j^r$ for all $j \in I$ such that

(i) if $r = 1$ and $q_i^1 = F_i^1(\theta_i, \tau_i^1, a_i)$ then $h_i^1(q_i^1) = \tau_i^1$ and (abusing notation) $a_i = f_i^*(\theta_i, q_i^1)$; and

(ii) if $r \geqslant 2$ and $q_i^r = F_i^r(\tau_i^r, q_i^{r-1})$ then $h_i^r(q_i^r) = \tau_i^r$ and $\phi_i^{r-1}(q_i^r) = q_i^{r-1}$.

We now construct a function F_i^{r+1}. To this effect fix any $(\bar{\tau}_i^{r+1}, \bar{q}_i^r) \in D_i^{r+1}$. If $r \geqslant 2$, define $\bar{\mu}_i^{r+1}$ to be the unique probability distribution over $\Theta_{-i} \times A_{-i} \times Q_{-i}^{r-1} \times T^r_{-i} \times Q^r_{-i}$ obtained as follows: First pick $(\theta_{-i}, a_{-i}, q_{-i}^{r-1})$ via $\bar{q}_i^r$; next, conditional on $(\theta_{-i}, a_{-i}, q_{-i}^{r-1})$ pick τ^r_{-i} via $\bar{\tau}_i^{r+1}(\cdot \mid \theta_{-i})$; and finally conditional on $(\theta_{-i}, a_{-i}, q_{-i}^{r-1}, \tau^r_{-i})$, pick $q^r_{-i} = F^r_{-i}(\tau^r_{-i}, q_{-i}^{r-1})$. If $r = 1$, $\bar{\mu}_i^{r+1}$ is constructed as above except that we eliminate all references to q_{-1}^{r-1} and Q_{-i}^{r-1} and replace $F^r_{-i}(\tau^r_{-i}, q_{-i}^{r-1})$ with $F^1_{-i}(\theta_i, \tau^r_{-i}, a_{-i})$. Define $\bar{q}_i^{r+1} \equiv F_i^{r+1}(\bar{\tau}_i^{r+1}, q_i^r)$ to be the marginal of $\bar{\mu}_i^{r+1}$ on $\Theta_{-i} \times A_{-i} \times Q^r_{-i}$. The claim below will imply that F_i^{r+1} maps $D_j^{r+1} \to Q_j^{r+1}$ and satisfies the induction hypotheses for $r+1$. From the hypotheses of Proposition 9.1 we have the existence of the function F_j^r for $r = 1$. Hence by induction such functions exist for all r.

Finally, for any $\tau_i = (\theta_i, \tau_i^1, \tau_i^2, \ldots) \in T_i$ and $a_i \in A_i$, define $F_i(\tau_i, a_i) = (\theta_i, q_i^1, q_i^2, \ldots)$ where $q_i^1 = F_i^1(\theta_i, \tau_i^1, a_i)$ and for $r \geqslant 2$, $q_i^r = F_i^r(\tau_i^r, q_i^{r-1})$. It should be clear that $F_i\colon T_i \times A_i \to Q_i$ and the conclusions of Lemma 9.1.1 hold. It should also be clear that we can replace A_i with the subset $\bar{A}_i$ everywhere in the proof to obtain Lemma 9.1.1 for $\{\bar{A}_i\}_{i \in I}$.

Claim. (a) $h_i^{r+1}(\bar{q}_i^{r+1}) = \bar{\tau}_i^{r+1}$ *and* (b) $\phi_i^r(\bar{q}_i^{r+1}) = \bar{q}_i^r$.

Proof. (a) Fix any $S \subseteq \Theta_{-i} \times T^r_{-i}$. Then for $r \geqslant 2$, (from the definition of h_i^{r+1}) $h_i^{r+1}(\bar{q}_i^{r+1})(S) = \bar{q}_i^{r+1}(\{(\theta_{-i}, h^r_{-i}(\bar{q}^r_{-i})) \in S\})$ (so from the definition of $\bar{q}_i^{r+1}$) $= \bar{\mu}_i^{r+1}(\{(\theta_{-i}, h^r_{-i}(F^r_{-i}(\tau^r_{-i}, q_{-i}^{r-1}))) \in S\})$ (and from the induction hypothesis on F^r_{-i}) $= \bar{\mu}_i^{r+1}(\{(\theta_{-i}, \tau^r_{-i}) \in S\})$ (so from the construction of $\bar{\mu}_i^{r+1}$) $= \bar{\tau}_i^{r+1}(S)$. Hence we have shown part (a) of the claim for $r \geqslant 2$. The situation for $r = 1$ is similar with obvious changes in notation. Part (b) follows in a manner similar to part (a). ∎

Proof of Proposition 9.1 (Cont'd). Define $\Omega_0 \equiv \Theta \times \prod_{n=1}^\infty A_n \times Z_n$, $\Omega_1 \equiv \prod_{n=1}^\infty T_n \times Q_n$ and $\Omega \equiv \Omega_0 \times \Omega_1$. We will first define for each $\omega_0 \in \Omega_0$ an element $\omega_1(\omega_0) \in \Omega_1$. We will then define μ_i to be the measure over Ω obtained by first choosing ω_0 via P_{AZ} then choosing ω_1 via $\omega_1(\omega_0)$. Clearly the marginal of μ_i on Ω_0 is P_{AZ} as required by the proposition. It only remains to show that we can construct $\omega_1(\omega_0) \in \Omega_1$ "consistently."

So fix an $\omega_0 = (\theta, a_1, z_1, a_2, z_2, \ldots) \in \Omega_0$. Given $\theta = \{\theta_i\}_{i \in I}$, pick for each $i \in I$ any $\tau_{i1} \in T_i$ which has an associated attribute vector θ_i. Given (τ_{i1}, a_{i1})

and referring to Lemma 9.1.1 define $q_{i1} = F_1(\tau_{i1}, a_{i1})$. We proceed by induction. Suppose for some N we have defined (τ_{in}, q_{in}) for $n = 1, ..., N$, with the property that $q_{in} = F_i(\tau_{in}, a_{in})$ for $n = 1, ..., N$. We will assume that the probability law of the observation process is common knowledge. Omitting the details (see [23]) this means we can talk about agent i's belief hierarchy over $\Theta_{-i} \times A_{-i}$ conditional on the observation of z_{iN}, $q_{iN}(\cdot \mid z_{iN}) \in Q_i$. Define $\tau_{iN+1} \equiv h_i(q_{iN}(\cdot \mid z_N))$ to be the induced belief hierarchy over θ and use Lemma 9.1.1 again and define $q_{iN+1} = F_i(\tau_{iN+1}, a_{iN+1})$. Hence by induction we obtain the sequence $\{q_{in}, a_{in}\}_{n=1}^{\infty}$. In particular we have the element $\omega_1(\omega_0) \in \Omega_1$ alluded to earlier. It should be clear that with this definition of $\omega_1(\omega_0)$ and the earlier defined μ_i for $i \in I$, μ_i is an ex ante subjective belief of agent i over Ω and that we obtain the conclusion of Proposition 9.1. ∎

Proof of Proposition 9.2. Define

$$L_n \equiv \ln v_n/(1 - v_n) - \ln v_1/(1 - v_1). \tag{9.2.1}$$

Define for each integer J, $B^j \equiv \{L_n \geqslant -J \text{ infinitely often}\}$, and $B \equiv U_{J=1}^{\infty} B^J$. Let $B^{\sim}$ and $B^{J\sim}$ denote the complements of the sets B and B^J, resp. Then $B^{\sim} = \bigcap_{J=1}^{\infty} B^{J\sim} = \{\lim_{n\to\infty} L_n = -\infty\}$. However, $L_n \to -\infty$ implies that $v_n \to 0$. Part (a) of the claim below then implies that $\mu^*(B) = 1$. Part (b) of the claim implies that on B, $\limsup_{n\to\infty} L_n = \infty$, which in turn implies that $\limsup_{n\to\infty} v_n = 1$ and hence from (9.1) that $\limsup_{n\to\infty} y_n = 1$. Very similar arguments imply that $\liminf_{n\to\infty} v_n = 0$ and $\liminf_{n\to\infty} y_n = 2/3$ with μ^* probability one. The claim therefore concludes the proof of Proposition 9.2.

Claim 1. *Outside of a set of sample paths with μ^*-probability zero, (a) v_n does not converge; and (b) for each integer J, on the set B^J,* $\limsup_{n\to\infty} L_n = \infty$.

Proof of Claim. First some preliminaries. Under the normality assumptions on $\{\varepsilon_n\}_{n=1}^{\infty}$, one may use Bayes rule to obtain

$$\begin{aligned} L_{n+1} &= L_n - (1/2)\{[p_{n+1} - (\alpha' - 1/3)]^2 - [p_{n+1} - (\alpha'' - 1/3)]^2\} \\ &= L_n + 5/54 - y_{n+1}/9 + \varepsilon_{n+1}/3. \end{aligned} \tag{9.2.2}$$

Fix any constant K, and define

$$d_N \equiv 3(K - L_N - 5/54 + y_{N+1}/9). \tag{9.2.3}$$

Define $\mathscr{F}_N^* \equiv \sigma(\{\varepsilon_n, y_n\}_{n=1}^{N}, y_{N+1})$ to be σ-field generated by the specified variables for any $N \geqslant 1$. Then (9.2.2) implies that $\text{Prob}(L_{N+1} \geqslant K \mid \mathscr{F}_N^*) = \text{Prob}(\varepsilon_{N+1} \geqslant d_N \mid \mathscr{F}_N^*)$, which, since ε_{N+1} is independent of $\mathscr{F}_N^*$, is equal to

$\text{Prob}(\varepsilon_{N+1} \geqslant d_N)$. From the "conditional" Borel-Cantelli lemma (see [9, p. 26]) we may conclude that on each sample path,

$$\sum_{N=1}^{\infty} \text{Prob}(\varepsilon_{N+1} \geqslant d_N) = \infty$$

$$\text{implies that} \quad \{L_{N+1} \geqslant K\} \text{ occurs for infinitely many } N. \tag{9.2.4}$$

(a) From (9.1) and the fact that agents are identical, to prove (a) it clearly suffices to prove that y_n does not converge. Fix any sample path and suppose, per absurdem, that on that path the aggregate output y_n converges to some value y_∞. Define, along that sample path, $\bar{p}_\infty \equiv \lim_{N \to \infty} (1/N) \sum_{n=1}^{N} p_n$. Denote by $f_N(p_1, ..., p_N \mid \alpha)$ the density or likelihood function of the ith agent's beliefs of $\{p_1, ..., p_N\}$ conditional on the intercept parameter α and define

$$M(\alpha) \equiv \lim_{n \to \infty} (1/N) \ln f_N(p_1, ..., p_N \mid \alpha)$$

and

$$A_\infty \equiv \{\alpha \in (\alpha', \alpha'') \mid M(\alpha) = \sup_{\theta \in (\alpha', \alpha'')} M(\theta)\}.$$

Then, using the strong law of Large numbers to set $(1/N) \sum_{n=1}^{N} \varepsilon_n = 0$, one may show that

$$\bar{p}_\infty = \alpha^* - y_\infty/3, \tag{9.2.5}$$

and

$$A_\infty = \inf_{\alpha \in (\alpha', \alpha'')} [\bar{p}_\infty - (\alpha - 1/3)]^2 + \text{terms independent of } \alpha. \tag{9.2.6}$$

Applying [32, Theorem 1], we may conclude that agent i's limiting beliefs will have support in the set A_∞, the "asymptotic carrier." We shall obtain a contradiction by showing that $A_\infty \neq \{\alpha'\}$, $A_\infty \neq \{\alpha''\}$ and $A_\infty \neq \{\alpha', \alpha''\}$.

First suppose $A_\infty = \{\alpha'\}$ on some sample path. Then $v_n \to 1$, so from (9.1), $y_{in} \to 1$ and from (9.2.5) $\bar{p}_\infty = 7/9$. Hence $[\bar{p}_\infty - (\alpha' - 1/3)]^2 = 4/81$ and $[\bar{p}_\infty - (\alpha'' - 1/3)]^2 = 1/81$ which is a contradiction to the fact that $A_\infty = \{\alpha'\}$. So $A_\infty \neq \{\alpha'\}$. Similar arguments show that $A_\infty \neq \{\alpha''\}$. Finally, suppose that $A_\infty = \{\alpha', \alpha''\}$. Then $[\bar{p}_\infty - (\alpha' - 1/3)]^2 = [\bar{p}_\infty - (\alpha'' - 1/3)]^2$ so $\bar{p}_\infty = 5/6$. Then (9.2.5) and (9.1) imply that

$$y_\infty = 5/6 \qquad \text{and} \qquad \lim_{n \to \infty} v_n = 1/2. \tag{9.2.7}$$

Define $\bar{L}_\infty \equiv -\ln v_1/((1 - v_1)$. Then from (9.2.7) and (9.2.1), $\lim_{N \to \infty} L_N = \bar{L}_\infty$. Fix any $\delta > 0$, define $K = \bar{L}_\infty + \delta$ and let d_N be as in (9.2.3). On any sample path where (9.2.7) holds, $\lim_{N\to\infty} d_N \to 3\delta$ so $\text{Prob}(\varepsilon_{N+1} \geqslant d_N) \to P(\varepsilon_1 \geqslant 3\delta) > 0$. We therefore conclude from (9.2.4) that on such a sample path $L_{N+1} \geqslant \bar{L}_\infty + \delta$ infinitely often. This contradicts the initial assertion that $\lim_{N\to\infty} L_N \to \bar{L}_\infty$. Hence $A_\infty \neq \{\alpha', \alpha''\}$.

(b) Fix any $K > 0$ and let d_N be as in (9.2.3). From (9.1), $y_n \leqslant 1$. Hence on the event $\{L_N \geqslant -J\}$, if we define $\bar{d} \equiv 3\{K + J - (5/54) + (1/9)\}$ then $d_N \leqslant \bar{d}$. So $\text{Prob}(\varepsilon_{N+1} \geqslant d_N) \geqslant \text{Prob}(\varepsilon_1 \geqslant \bar{d}) > 0$. Since the event $\{L_N \geqslant -J\}$ occurs infinitely often on B^J, (9.2.4) implies $\{L_{N+1} \geqslant K\}$ occurs infinitely often. Since K is arbitrary this implies part (b) of the claim. ▮

REFERENCES

1. W. Ambruster and W. Boge, Bayesian game theory, *in* “Game Theory and Related Topics” (O. Moeschlin and D. Pallachke, Eds.), pp. 17–28, North-Holland, Amsterdam, 1979.
2. C. Berge, “Topological Spaces,” Oliver and Boyd, Edinburgh, 1963.
3. B. D. Bernheim, Rationalizable strategic behavior, *Econometrica* **52**, No. 4 (1984), 1007–1028.
4. P. Billingsley, “Convergence of Probability Measures,” Wiley, New York, 1968.
5. D. Blackwell and L. Dubins, Merging of opinions with increasing information, *Ann. Math. Statist.* **38** (1963), 882–886.
6. L. Blume and D. Easley, Rational expectations equilibrium an alternative approach, *J. Econ. Theory* **34** (1984), 116-129.
7. A. Brandenburger, Information and learning in market games (mimeo), Graduate School of Business, Stanford University, 1984.
8. M. Bray, Learning, estimation and the stability of rational expectations, *J. Econ. Theory* **26** (1982), 318–319.
9. Y. S. Chow, H. Robbins, and D. Siegmund, “Great Expectations The Theory of Optimal Stopping,” Houghton Mifflin Co., Boston, 1971.
10. K. L. Chung, “A Course in Probability Theory,” Academic Press, New York, 1974.
11. M. Feldman, An example of convergence to rational expectations with heterogeneous beliefs, *International Econ. Rev.* **28** (1988), 635–650.
12. R. Frydman, Towards an understanding of market processes individual expectations and convergence to rational expectations, *Amer. Econ. Rev.* **72** (1982), 652–668.
13. R. Guesnerie, An exploration of the eductive justifications of the rational-expectations hypothesis, *Amer. Econ. Rev.* (1992), 1254–1278.
14. J. C. Harsanyi, Games with incomplete information played by Bayesian players, parts I, II, III, *Management Sci.* **14**, Nos. 3, 5, 7 (1967, 1968).
15. J. S. Jordan, Bayesian learning in normal form games, *Games Econ. Behav.* **3**, No. 1 (1991), 60–81.
16. J. S. Jordan, Bayesian learning in repeated games, *Games Econ. Behav.* **9**, No. 1 (1995), 8–20.
17. E. Kalai and E. Lehrer, Rational learning leads to Nash equilibrium, *Econometrica* **61**, No. 5 (1993), 1019–1047.
18. M. Keynes, “The General Theory of Employment, Interest and Money,” Harcourt Brace Jovanovich, New York, 1964.

19. A. Marcet and T. Sargent, Convergence of least squares learning mechanisms in self referential linear stochastic models, *J. Econ. Theory* **48** (1989), 337–368.
20. J.-F. Mertens, and S. Zamir, Formalization of Bayesian analysis for games with incomplete information, *Int. J. Game Theory* (1985), 141–129.
21. Y. Nyarko, Bayesian learning without common priors and convergence to Nash equilibria, forthcoming, *Econ. Theory*, 1997.
22. Y. Nyarko, Bayesian learning in leads to correlated equilibria in normal form games, *Econ. Theory* **4**, No. 6 (1994), 821–841.
23. Y. Nyarko, The Savage–Bayesian foundations of economic dynamics, manuscript, New York University, 1993.
24. Y. Nyarko, The 'types' of a Bayesian equilibrium, manuscript, New York University, 1993.
25. Y. Nyarko, "Bayesian Learning Without Common Priors and Convergence to Nash Equilibrium," C. V. Starr Working Paper No. 92–25, New York University, 1992.
26. Y. Nyarko, The number of equations versus the number of unknowns the convergence of Bayesian posterior processes, *J. Econ. Dynamics Control* **15** (1991), 687–713.
27. Y. Nyarko, "Bayesian Rationality and Learning without Common Priors," C.V. Starr Working Paper No. 90–45, New York University, 1990.
28. K. P. Parthasarathy, "Probability Measures on Metric Spaces," Academic Press, New York, 1967.
29. D. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52**, No. 4 (1984), 1029–1050.
30. L. J. Savage, "The Foundations of Statistics," Wiley, New York, 1954.
31. R. Townsend, Market anticipations, rational expectations, and Bayesian analysis, *Int. Econ. Rev.* **19**, No. 2 (1978), 481–494.
32. K. Yamada, Asymptotic behavior of posterior distributions for random processes under incorrect models, *J. Math. Anal. Apply.* **56** (1976), 294–308.