ECONOMIC RESEARCH REPORTS

THE SAVAGE-BAYESIAN FOUNDATIONS OF ECONOMIC DYNAMICS

BY

Yaw Nyarko

RR # 93-35 *August, 1993*

C. V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

The Savage-Bayesian Foundations of Economic Dynamics

Yaw Nyarko[1]
New York University
269 Mercer St. Rm 723
New York N.Y. 10003, U.S.A.
Spring 1993

[1] I thank the C.V. Starr Center and the Presidential Fellowship Fund at New York University for their generosity. I also thank Professor Jim Jordan for many helpful discussion. All errors in this paper are my own responsibility.

The Savage-Bayesian Foundations of Economic Dynamics

Yaw Nyarko

Abstract

In this paper I provide a framework general enough to study economic dynamics in a multi-agent model where each agent has imperfect information both on the "fundamentals" of the economy (or "game") and also imperfect information on the actions or strategies being used by other agents. This paper generalizes the work of Ambruster and Boge (1979), Boge and Eisele (1979), Mertens and Zamir (1985), and others in the following ways: **(i)** First, the emphasis of this paper will be on ***dynamic*** models. **(ii)** Like Harsanyi (1968), this paper will "go behind the veil" and model an ***ex ante*** period before agents are "born" and realize their individual characteristics. Agents will be assumed to have ex ante subjective beliefs, which do ***not*** necessarily obey the common prior assumption. **(iii)** The study of a dynamic model leads naturally to a new concept of ("versions" or equivalence classes of) ***belief hierarchies over a random variable*** at some future date n conditional on data observed by that date. **(iv)** I consider the basic space of uncertainty to be both the fundamentals ***and*** the strategies of the agents. This allows for a unification of the literature (where different underlying spaces are used) and also allows for a precise definition of the various notions of a "type" used in the literature.

Journal of Economic Literature No's: C70, C73, D81, D82, D83, D84.

Mailing Address
Yaw Nyarko
New York University
269 Mercer St. Rm 723
New York N.Y. 10003, U.S.A.

1. Introduction

1.1. In this paper I provide a framework general enough to study economic dynamics in a multi-agent model where each agent has imperfect information both on the "fundamentals" of the economy (or "game") and also imperfect information on the actions or strategies being used by other agents. The framework will be "Savage-Bayesian" in the sense that agents' preferences obey the axioms of Savage (1954). In particular agents form probability beliefs over ***all*** items they are uncertain about and maximize their subjective expected utility. Imperfect information in multi-agent models gives rise to belief-hierarchies of agents (i.e., an agent's belief about the other agents' beliefs, an agent's belief about the beliefs other agents have about the beliefs of other agents, etc.). Modelling such multi-agent economies or games has been studied by many authors (e.g., Ambruster and Boge (1979), Boge and Eisele (1979), Mertens and Zamir (1985), Tan and Werlang (1988), Brandenberger and Dekel (1993) and Heifetz (1990)).

A principal motivation for this paper is that here, unlike the earlier-mentioned papers, I will provide a framework rich enough to handle the recent work on "learning" and on "Bayesian learning" in particular (as for example in Townsend (1978), Feldman (1988), Jordan (1991), Kalai and Lehrer (1991) and Nyarko (1991b, 1992 and 1993)). In particular, this paper deals with the following four issues which, to the best of my knowledge, have not been addressed in the literature: **(i)** First, the emphasis of this paper will be on ***dynamic*** models where at each date actions are taken and agents make observations. The introduction of dynamics requires careful attention to probability zero issues and the regularity of conditional probabilities. **(ii)** Like Harsanyi (1968), this paper will "go behind the veil" and model an ***ex ante*** period before agents are "born" and realize their individual characteristics. Agents will be assumed to have ex ante subjective beliefs. It will be shown that these ex ante beliefs are in many ways without loss of generality. Their use however allows us to get around many technical problems. Unlike Harsanyi (1968) however, I do not use the common prior assumption (which is used with a great loss of generality -see Nyarko (1991a))! I replace this with mutual absolute continuity assumptions on the ex ante beliefs of agents. **(iii)** The study of a dynamic model leads naturally to the question of ***belief hierarchies over a random variable*** at some future date n conditional on data observed by that date. These are defined in this paper. I show that it is necessary to talk about ***equivalence*** classes of these hierarchies (just as we have equivalence classes, or versions, of conditional probabilities). **(iv)** Some papers in the literature consider the basic space of uncertainty to be the space of "fundamentals" like utility parameters (as e.g., in Mertens and Zamir (1985)) while others consider it to be the space of actions. Following Boge and Eisele (1979), I consider the basic space of uncertainty to be both the fundamentals ***and*** the strategies of the agents. In defining notions of a "type" no distinction is made in the literature as to the underlying space of uncertainty. The work of this paper is continued in a sequel, Nyarko (1993b) where the construction of this paper is used. In that paper I distinguish between various notions of a ***type.*** For example, a "Savage-Bayesian" type is a belief hierarchy over the fundamentals and actions. This is the notion of a type required in multi-agent decision theory. A second notion of a type is the "Harsanyi" type. This is the induced belief hierarchy over only the fundamentals. Harsanyi types are required in the definition of a Harsanyi Bayesian Nash ***equilibrium.*** The framework of this paper allows for a discussion of these two (and other) notions of a type.

Like Brandenberger and Dekel (1993) and Heifetz (1990) our underlying spaces will ***not*** be assumed to be compact. Instead we suppose they are complete and separable metric spaces. This paper will use standard arguments and constructions of probability theory.

1.2. A Motivating Example. The following model of competitive firms facing an unknown demand curve is studied in much greater detail in Nyarko (1991b): 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 $p_n = \alpha - \beta y_n + \epsilon_n$, 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. We suppose that the parameter β of the demand curve is "common knowledge" among the agents. However there is imperfect information over the parameter α. 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.

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, $G_n x \equiv \int_0^1 (E_{in}x) di$. If agents do not know the beliefs of others then in general there will be uncertainty over expressions like $E_{in}G_n x$, agent i's expectation of the average opinion of x, and $G_n^2 x$, the average opinion of the average opinion of x. Inductively, we may defined $G_n^r x$ to be the r-times average opinion of the average opinion ... of x. ***Maximizing behavior*** of firms (which we write as ***(MB)***) implies the following:

(MB) $\quad y_{in}=E_{in}\alpha-\beta E_{in}y_n$ so by integration over i, $\quad y_n=G_n\alpha-\beta G_n y_n$.

If there is respectively 1-level knowledge of (MB) (i.e., if agents know that other agents engage in (MB)); or 2-level knowledge of (MB) (i.e.,. if agents know that other agents know that other agents engage in (MB)) or R-levels of knowledge of (MB) we obtain:

(1-level knowledge of (MB)): $y_{in}=E_{in}\alpha-\beta E_{in}G_n\alpha+\beta^2E_{in}G_ny_n$.

(2-level knowledge of (MB)): $y_{in}=E_{in}\alpha-\beta E_{in}G_n\alpha+\beta^2E_{in}G_n{}^2\alpha-\beta^3E_{in}G_n{}^2y_n$.

(R-level knowledge of (MB)): $y_{in}=\Sigma_{r=1}{}^{R+1}(-\beta)^{r-1}E_{in}(G_n{}^{r-1}\alpha)+(-\beta)^{R+1}E_{in}G_n{}^Ry_n$(where $G^0\alpha=\alpha$).

We suppose that the above decision-problem occurs at each date n=1,2,... . The goal of this paper is to provide a probabilistic framework general enough to formally model all of the above, including the expectations which, recall, are conditional upon the information agents observe by the beginning of date n.

2. Some Terminology and Mathematical Preliminaries

2.1. I is the ***finite*** set of economic agents. Nature is agent 0, and is not a member of I. Given any collection of sets $\{X_i\}_{i\epsilon I}$, we define $X\equiv\Pi_{i\epsilon I}X_i$ and $X_{-i}\equiv\Pi_{j\neq i}X_j$ unless otherwise stated; (given X_0 and $\{X_i\}_{i\epsilon I}$, we shall sometimes state that $X_{-i}\equiv X_0x\Pi_{j\neq i}X_j$). Given any collection of functions $f_i:X_i\rightarrow Y_i$ for $i\epsilon I$, $f_{-i}:X_{-i}\rightarrow Y_{-i}$ is defined by $f_{-i}(x_{-i})\equiv\Pi_{j\neq i}f_j(x_j)$. The cartesian product of metric spaces will always be endowed with the product topology. Let X be any metric space. $\mathcal{P}(X)$ 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 $\mathcal{P}(X)$ will be endowed with the weak topology of measures; (see Billingsley (1968) for more on this). The following fact will be used repeatedly: If X is a complete and separable metric space then so is $\mathcal{P}(X)$. (See, e.g., Parthasarathy Theorems II.6.2 and II.6.5.) 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; (see appendix A for definitions).

2.2. A Marginal ⊗ A Conditional = A Joint Probability. Let X and W be two complete and separable metric spaces. Suppose we are given a ("marginal") distribution, Ψ', over X; i.e., $\Psi' \epsilon \mathcal{P}(X)$. Let $G: X \rightarrow \mathcal{P}(W)$ be any function mapping X into the set of probability measures on W. Let G(.;x) denote the value of G at x (so $G(.;x) \epsilon \mathcal{P}(W)$). Then each x defines a probability G(.;x) on W ("conditional" on x). We may therefore "integrate" the conditionals with respect to the marginal to obtain a joint distribution, Ψ, over XxW. This joint probability, Ψ, will have a marginal over X equal to Ψ' and a conditional over W given x equal to G(.;x). We shall use the notation $\Psi' \otimes G$ or $\Psi' \otimes G(.;x)$ to denote this joint probability and refer to it as the "***product***" of Ψ' and G(.;x). (For this "product" operation we will require the measurability of G(.;x) in x; in the appendix this is shown to be equivalent to the requirement that G(.;x) be a regular conditional probability.) In the lemma below we formalize this discussion and show that the "product" operation is well-defined. The proof of this and all other results appear in the appendix.

Lemma 2.3. Let W, X, Ψ' and G be as above. For each (Borel-measurable) subset $S \subseteq XxW$, define $S_x \equiv \{w \epsilon W : (x,w) \epsilon S\}$ and $\Psi(S) \equiv \int_X G(S_x;x) d\Psi'$. Then

i. Ψ is a well-defined a probability measure over XxW;

ii. the marginal of Ψ on X is Ψ'; i.e., for all (measurable) $D' \subseteq X$, $\Psi(\{x \epsilon D'\}) = \Psi'(D')$; and

iii. $G(.;x)$ is a version of the conditional probability of Ψ given x; i.e., for all (measurable) $S \subseteq XxW$, if $\Psi(S \mid x)$ is any version of the conditional probability of S conditional on x then $\Psi(S \mid x) = G(S_x;x)$ for Ψ' almost every x.

3. Hierarchies of Beliefs and the Space $B_i(Y)$

3.1. Recall that I is the set of agents and nature is referred to as agent 0 (not a member of I). Suppose we are given a collection of complete and separable metric spaces Y_0 and $\{Y_i\}_{i\epsilon I}$. We shall consider Y_i to be the set pertaining to agent i; this will have the meaning that i "knows" its own value of $y_i \epsilon Y_i$. We consider Y_0 to be the parameters of "nature." We proceed to construct the space of hierarchies of beliefs over the space $Y = Y_0 x \Pi_{i\epsilon I} Y_i$. Construct the sets $\{B_i^r\}^{\infty}_{r=1}$ inductively as follows:

$$B_i^1 \equiv \mathcal{P}(Y_{-i}) \qquad \text{where } Y_{-i} \equiv Y_0 x \Pi_{j \neq i} Y_j; \tag{3.2}$$

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

$$B_i^{r+1} \equiv \mathcal{P}(B^r_{-i} x Y_{-i}). \tag{3.3}$$

An element $b_i^1 \epsilon B_i^1$ represents agent i's belief about $y_{-i} \epsilon Y_{-i}$ and shall be referred to as agent i's first order belief. An element $b_i^2 \epsilon B_i^2$ 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 $b_i^r \epsilon B_i^r$ is i's r-th 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 Y_{-i}. To express this relation we define the functions $\phi_i^r: B_i^{r+1} \rightarrow B_i^r$ inductively as follows: For any subset $D \subseteq Y_{-i}$,

$$\phi_i^1(b_i^2)(D) \equiv b_i^2(\{B^1_{-i} x D\}) \text{ for all } b_i^2 \epsilon B_i^2; \tag{3.4}$$

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

$$\phi_i^r(b_i^{r+1})(D) \equiv b_i^{r+1}(\{(b_{-i}^r, y_{-i}) \epsilon B_{-i}^r x Y_{-i}: (\phi_{-i}^{r-1}(b_{-i}^r), y_{-i}) \epsilon D\}). \tag{3.5}$$

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

$$B_i \equiv \{(b_i^1, b_i^2, \ldots) \epsilon \Pi^{\infty}_{r=1} B_i^r: b_i^r = \phi_i^r(b_i^{r+1}) \text{ for all } r \geq 1\}. \tag{3.6}$$

We must stress that the set B_i is defined by the underlying space of uncertainty Y. The property in (3.6) that $b_i^r = \phi_i^r(b_i^{r+1})$ for all $r \geq 1$ shall be called the ***probabilistic coherence property*** of belief hierarchies. This requires that lower order beliefs be a "projection" of the higher order beliefs.

3.7. Remark. The construction above differs in one respect from that of Mertens and Zamir (1985). An agent's belief of any order is a belief about other agents and does not include a belief about the agent herself. For example $B_i^1 = \mathcal{P}(Y_{-i})$ as opposed to $\mathcal{P}(Y)$ and $B_i^2 = \mathcal{P}(B_{-i}^1 x Y_{-i})$ as opposed to say $\mathcal{P}(B^1 x Y)$. In this regard the construction is the same as Myerson (1985). This difference will be important later when we will need to specify beliefs of agents ***before*** they choose their own strategies. We do not want to have the beliefs specifying strategies that have not yet been chosen!

One may be curious why we did not construct the types sets by defining $B_i^{r+1}=\mathcal{P}(B^r_{-i})$ as opposed to $\mathcal{P}(B^r_{-i}xY_{-i})$. The reason is that we seek to allow the i-th agent to have beliefs under which various orders of beliefs of other agents are correlated. For example the i-th agent may think that if $y=\bar{y}$ then other agents have first order beliefs $\bar{b}_{-i}^1$ while if $y=\hat{y}$ then the other agents have first order beliefs $\hat{b}_{-i}^1$. The above construction allows for these correlations. One may also wonder why we did not construct the r-th order belief to be the joint probability over <u>all</u> lower order beliefs. In particular, one may ask why we did not define $B_i^{r+1}= \mathcal{P}(Y_{-i}xB_{-i}^1xB_{-i}^2x\ldots xB_{-i}^r)$. This is because the r-th order belief determines all lower order beliefs via a relation similar to the equality in (3.6); hence the inclusion of lower order beliefs is a redundancy.

3.8. An Interpretation of the Probabilistic Coherence property. We proceed to show that a role of the probabilistic coherence property is the following: It ensures that for each agent i, any given set D (of lower order beliefs of others) is assigned the same probability by any two higher order beliefs of that agent. (This is very similar to the role played by the Kolmogorov consistency conditions in defining an infinite dimensional probability from its finite dimensional distributions.) So fix any two integers r and r' with $r<r'$. Let ϕ_i^r be as in (3.5). For all $i\epsilon I$, define $\Phi_i^{r,r'}:B_i^{r'}\rightarrow B_i^r$ to be the "projection" from the higher dimensional set $B_i^{r'}$ to the lower dimensional set B_i^r as follows:

$$\Phi_i^{r,r'}(b_i^{r'}) \equiv \phi_i^r(\phi_i^{r+1}(\ldots(\phi_i^{r'-1}(b_i^{r'})))). \tag{3.9}$$

If $b_i=(b_i^1,b_i^2,\ldots)\epsilon\Pi^\infty_{r=1}B_i^r$ obeys the probabilistic coherence property then of course $b_i^r=\Phi_i^{r,r'}(b_i^{r'})$.

When $r'=r+1$ then $\Phi_i^{r,r'}=\phi_i^r$. Fix any subset $D\subseteq B_{-i}^{r-1}xY_{-i}$. For $\rho=0,1,2,..$, define $D^{+\rho}$ to be the set D "uplifted by ρ coordinates" onto $B_{-i}^{r-1+\rho}xY_{-i}$; i.e., define

$$D^{+\rho} = \{(b_{-i}^{r-1+\rho},y_{-i})\epsilon B_{-i}^{r-1+\rho}xY_{-i} \text{ such that } (\Phi_{-i}^{r-1,\,r-1+\rho}(b_{-i}^{r-1+\rho})),y_{-i})\epsilon D\} \tag{3.10}$$

and $D^{+0}\equiv D$. Fix any $b_i=(b_i^1,b_i^2,\ldots)\ \epsilon\ B_i$ and $D\subseteq B_{-i}^{r-1}xY_{-i}$. Then $b_i^r(D)$ is the probability assigned by b_i^r to the set D; and $b_i^{r'}(D^{+(r'-r)})$ is the probability assigned by $b_i^{r'}$ to the set D after it has been "uplifted" by r'-r coordinates onto $B_{-i}^{r'-1}xY_{-i}$ (so that it is on the appropriate domain). $b_i^r(D)$ and $b_i^{r'}(D^{+(r'-r)})$ are two ways of measuring the probability of the set D. If b_i^r and $b_i^{r'}$ are to represent the beliefs of the same agent, these two ways of measuring the set D should be equal. The lemma below shows if $b_i\epsilon B_i$ (so that the probabilistic coherence property holds) then this is indeed the case.

Lemma 3.11. Fix any $b_i=(b_i^1,b_i^2,\ldots)\ \epsilon\ B_i$, any integers $r\geq 1$ and $\rho\geq 1$, and any $D\subseteq B_{-i}^{r-1}xY_{-i}$. Let $D^{+\rho}$ be as in (3.10). Then $b_i^r(D)=b_i^{r+\rho}(D^{+\rho})$.

4. The Mapping P_i: $B_i \rightarrow \mathcal{P}(B_{-i}xY_{-i})$

4.1. Fix any element $b_i=(b_i^1,b_i^2,\ldots)\ \epsilon B_i$. We will now construct an "associated" probability $P_i(b_i)$ $\epsilon\ \mathcal{P}(B_{-i}xY_{-i})$ with the property that for each integer r, the marginal of $P_i(b_i)$ on $B_{-i}^rxY_{-i}$ is equal to b_i^{r+1}. Recall that by definition B_{-i} is a subset of the infinite cartesian product $\Pi^\infty_{r=1} B_{-i}^r$. Define $PROJ_{-i,r}$: $B_{-i}xY_{-i} \rightarrow B_{-i}^rxY_{-i}$ to be the projection of the space $B_{-i}xY_{-i}$ onto $B_{-i}^rxY_{-i}$. The inverse of the projection mapping, $(PROJ_{-i,r})^{-1} D^r$, is the "upliftment" of the subset $D^r \subseteq B_{-i}^r xY_{-i}$ to the infinite cartesian product $B_{-i}xY_{-i}$. Fix any integer r and define the class Δ^r of r-cylinder subsets of $B_{-i} xY_{-i}$

to be those which are "upliftments" of some subset of $B_{-i}^r x Y_{-i}$; i.e.,

$$\Delta^r \equiv \{D \subseteq B_{-i} x Y_{-i}\text{: } D = (PROJ_{-i\,,\,r})^{-1} D^r \text{ for some (measurable) } D^r \subseteq B_{-i}^r x Y_{-i}\}. \qquad (4.2)$$

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

$$P_i(b_i)(D) = b_i^{r+1}(\{PROJ_{-i\,,\,r}\, D\}). \qquad (4.3)$$

From the probabilistic coherence condition implicit in the definition of B_i, and in particular from the "projection" result of Lemma 3.11 it is easy to check that for each $b_i \epsilon B_i$, $P_i(b_i)$ in (4.3) is "well-defined" over $\bigcup^{\infty}_{r=1} \Delta^r$ in the following sense: if a set D lies in both Δ^r and $\Delta^{r'}$ for any integers r and r′ then the definition of $P_i(b_i)(D)$ in (4.3) results in the same answer. The relation (4.3) defines for each $b_i \epsilon B_i$ a probability measure $P_i(b_i)$ on the set of cylinder sets, $\bigcup^{\infty}_{r=1} \Delta^r$, of $B_{-i} x Y_{-i}$. The "well-definedness" property just mentioned (together with the probabilistic coherence condition) implies what is referred to in standard probability texts as the "consistency condition". We may thererfore apply a standard probability extension argument to show that the probability $P_i(b_i)$ extends to a unique probability measure over all Borel subsets of $B_{-i} x Y_{-i}$. (See for example Parthasarathy (1967), Theorem 4.2., p. 143).) Since cylinder sets are "convergence determining" it is easy to check that the mapping $P_i: B_i \rightarrow \mathcal{P}(B_{-i} x Y_{-i})$ is ***continuous***.

4.4. Remark. Brandenberger and Dekel (1993) also construct the function $P_i(b_i)$ in a similar manner to that above. Our formulation differs slightly from theirs in many of the ways discussed in (3.7). For this reason and because this construction is so fundamental to our analysis we have repeated parts of the argument here. Brandenberger and Dekel (1993) go on to show that the spaces B_i and $\mathcal{P}(B_{-i} x Y_{-i})$ are homomorphic and discuss some common knowledge questions associated with the construction.

5. The Economic Model

5.1. Time is discrete and has dates n=1,2,3,... . At each date n agent i chooses an action a_{in} in an action space A_i. Let $z_{in} \in Z_i$ denote the vector of all observations of agent i during the course of date n. We assume that agents observe their own actions, so z_{in} is a vector which includes a specification of a_{in}. If all agents observe the same information then of course $z_{in} = z_{jn}$ for all i and j in I. Since each z_{in} itself may be a very large vector this formulation allows agents to observe some common signals (e.g., market prices) as well as private signals. Just before choosing the date n action $a_{in} \in A_i$, agent i would have information on the date n partial history $z_i^{N-1} = \{z_{i1},\ldots,z_{iN-1}\} \in Z_i^{N-1}$. ($z_i^0$ is the null or empty history.) We suppose Z_i and A_i are complete and separable metric spaces for all $i \in I$. Define

$$F_{iN} \equiv \{f_{iN}: Z_i^{N-1} \rightarrow A_i \text{ with } f_{iN} \text{ Borel-measurable}\}, \quad F_i \equiv \Pi^{\infty}_{N=1} F_{iN} \text{ and } F \equiv \Pi_{i \in I} F_i. \tag{5.2}$$

A ***behavior strategy*** for agent i is any $f_i \in F_i$. We now require the following:

5.3. Assumption: For all i and N, the space of date N behavior strategies, F_{iN}, is endowed with a metric which makes it ***a complete and separable metric space.***

5.4. Remark. When working with "macro-economic" types of models like that of example 1.2, one may show (and hence assume) that the optimal actions of agents lie in some pre-specified compact set and are continuous functions of the history. In this case we may assume F_{iN} is the space of continuous functions endowed with the sup norm. In repeated games problems we may assume the action space is finite in which case we may endow F_{iN} with the topology of pointwise convergence. In either case F_{iN} will be a complete and separable metric space. We stress that precise nature of the metric is unimportant. It is the fact that it is a metric space which is important. At this stage all we need is enough structure to be able to talk about probability measures. Since we will be performing integration exercises, and in particular the "integration" in (2.2), we require the metric space to be complete and separable.

5.5. Attribute Vectors. We let $\Theta \equiv \Theta_0 x \Pi_{i\epsilon I} \Theta_i$ denote the space of ***"fundamentals"*** or ***attribute vectors*** of agents. $\theta_0 \epsilon \Theta_0$ will denote nature's attribute vector; this parameter will determine any underlying randomness of the economy. $\theta_i \epsilon \Theta_i$ denotes the utility parameter or attribute vector of agent i. Agent i's utility or payoff function is some function $u_i:\Theta_0 x \Theta_i x F_i x F_{-i} \to \mathbf{R}$ which depends upon nature's attribute vector, θ_0, agent i's attribute vector, θ_i, agent i's strategy vector, f_i, and the strategy vector of the other agents, f_{-i}. It will be assumed that the functional forms of the utility functions are common knowledge; however agents will have imperfect information over the attribute vectors and the behavior strategies of other agents. Θ_0 and Θ_i for each $i\epsilon I$ are assumed to be complete and separable metric spaces.

5.6. The Measure P_γ (or the "laws of economics") Define $\Gamma \equiv \Theta_0 x F$, the cartesian product of nature's attribute vector and the behavior strategies of agents, $F \equiv \Pi_{i\epsilon I} F_i$. We shall suppose that sequence of observations and actions of the economy, $\{z_n\}^{\infty}_{n=1} \in Z^{\infty}$ has a probability distribution P_γ which depends upon the true date 0 vector $\gamma=(\theta_0,\{f_i\}_{i\in I})\in\Gamma$. We may without loss of generality suppose that this probability distribution ***as a function of γ*** is "common knowledge" among the agents in the economy. Indeed, once we have specified γ we have specified all the elements which could possibly affect the evolution of the economy over time: The vector of Nature's attributes θ_0 determines all exogenous uncertainty and random variables while f_i determines the behavior of agent i. Hence P_γ should be considered to be the "definition" of γ. If P_γ is not "common knowledge" then we have not specified either the fundamentals or the strategies vectors of agents appropriately. Although P_γ is common knowledge as a function of γ, the value of γ itself will in general be unknown to agents within the economy. One may think of P_γ as the ***"laws of economics"*** since it is really the exogenously determined law of evolution of actions and observations in the economy as a function of the date 0 value of $\gamma\in\Theta_0 x F$. We shall assume that P_γ is a ***regular conditional probability*** on Z^{∞}. (Of course, by assuming that θ_0 includes a specification of the utility parameters of ***all*** agents, we may model the situation where agent i's utility function and P_γ are functions of agent j's attribute vector (for some or all $j\in I$), as in some formulations of adverse selection models in economics).

6. The Beliefs of Agents

6.1. Savage-Bayesian Types. At date 0 there is imperfect information over space of attribute vectors or "fundamentals," $\Theta=\Theta_0 x\Pi_{i\in I}\Theta_i$, ***and*** over the space of behavior strategies, $F=\Pi_{j\in I}F_j$. Agent i has imperfect information over the set $\Theta_{-i} x F_{-i}$ (where $\Theta_{-i}\equiv\Theta_0 x\Pi_{j\neq i}\Theta_j$). Let Q_i^{∞} be agent i's space of belief hierarchies over $\Theta x F$ defined and constructed as in section 3.1. (In that construction set $Y_0=\Theta_0$ and $Y_j=\Theta_j x F_j$ for all $j\epsilon I$; what we refer to here as Q_i^∞ is the same as what was referred to in that construction as B_i.) Any $q_i^\infty=(q_i^1,q_i^2,\ldots)\epsilon Q_i^\infty$ is a possible belief hierarchy for agent i over $\Theta x F$. At date 0 each agent i will be characterized by some attribute vector, $\theta_i \epsilon \Theta_i$, and some belief hierarchy $q_i^\infty \epsilon Q_i^\infty$. We refer to the tuple $q_i=(\theta_i,q_i^\infty)$ as agent i's ***Savage-Bayesian type*** and we define $Q_i\equiv\Theta_i x Q_i^\infty$ and $Q\equiv\Pi_{i\epsilon I}Q_i$. An agent's Savage-Bayesian type, $q_i=(\theta_i,q_i^\infty)$, contains all the information for that agent to engage in decision-making: preferences are specified by θ_i and beliefs specified by q_i^∞.

6.2. Behavior Strategy Choice Rules and Expected Utility Maximization.

We define a ***behavior strategy choice rule*** to be any (measurable) function $\nu_i''':Q_i\rightarrow\mathcal{P}(F_i)$ which determines agent i's (possibly randomized) behavior strategy as a function of that agent's Savage-Bayesian type, q_i. Define $U_i(q_i,f_i)$ to be the expected utility function of agent i of Savage-Bayesian type $q_i=(\theta_i,q_i^1,q_i^2,q_i^3,\ldots)$ obtained by integrating out the coordinates Θ_0 and F_{-i} from the utility function u_i of (5.5) with respect to the measure q_i^1:

$$U_i(q_i,f_i) \equiv \int_{\Theta_0 x F_{-i}} u_i(\theta_0,\theta_i,f_i,f_{-i})dq_i^1. \tag{6.3}$$

Conditional on any q_i an expected utility maximizer will choose a behavior strategy f_i^* to maximize the expression in (6.3). If there is more than one solution to this maximization problem the agent could in general randomize over the set of maximizers. Expected utility maximization will therefore under fairly general conditions result in a behavior strategy choice rule.

6.4. The State Space Ω.

We define the state space to be the set $\Omega\equiv Q x\Theta_0 x F x Z^\infty$. Any

$\omega=(\{q_i\}_{i\epsilon I},\theta_0,\{f_i\}_{i\epsilon I},z^\infty)\epsilon\Omega$ specifies the Savage-Bayesian types of agents, $\{q_i\}_{i\epsilon I}$, nature's attribute vector, θ_0, the vector of agents' behavior strategies, $\{f_i\}_{i\epsilon I}$, and the sample path of actions and observations, $z^\infty\epsilon Z^\infty$.

6.5. Agent i's belief $\mu_i(.;q_i)$ over Ω. We now ***assume*** the existence of a regular conditional probability $\mu_i'''(.;q_i)\ \epsilon\ \mathcal{P}(F_i)$ which determines the (perhaps randomized) behavior strategy of agent i conditional on any $q_i\epsilon Q_i$. This could come from expected utility maximization as in (6.3). Alternatively, it could be the result of some rule of thumb, for example. We however stress now that in ***modelling*** our agent, we do not require expected utility maximization; we merely require the existence of ***some*** behavior strategy choice rule $\mu'''(.;q_i)$.

We proceed to model the belief of agent i with Savage-Bayesian type q_i. Recall from section 4 that any hierarchy of beliefs, $q_i^\infty\epsilon Q_i^\infty$, induces a probability measure $P_i(q_i^\infty)$ over belief hierarchies of others. In particular, $P_i(q_i^\infty)$ is a probability over the space $Q_{-i}^\infty x\Theta_{-i}xF_{-i}$. Since $Q_j\equiv\Theta_jxQ_j^\infty$ and $\Theta\equiv\Theta_0x\Pi_{j\epsilon I}\Theta_j$, the space $Q_{-i}^\infty x\Theta_{-i}xF_{-i}$ is the same as $Q_{-i}x\Theta_0xF_{-i}$, so we may consider $P_i(q_i^\infty)$ a probability over the latter space. Let $1_i\ (q_i)$ be the probability measure over Q_i which assigns probability one to the given vector q_i; consider this to be a "marginal" on Q_i. $\mu_i'''(.;\hat{q}_i)$ is the strategy choice rule of agent i; consider this to be a "conditional" over F_i given ***any*** $\hat{q}_i\epsilon Q_i$. The product (as in (2.2)) of the marginal and the conditional, $1_i(q_i)\otimes\mu_i'''$, is a joint probability over Q_ixF_i. We suppose agent i knows her own Savage-Bayesian type and her behavior strategy choice rule, so this joint probability represents that agent's belief over Q_ixF_i. Now consider this probability as a "marginal" probability over Q_ixF_i and take the product (as in (2.2)) of this first with $P_i(.)$ (which, recall, is the "conditional" probability function equal to $P_i(\bar{q}_i^\infty)$, an element of $\mathcal{P}(Q_{-i}x\Theta_0xF_{-i})$, for ***any*** $\bar{q}_i=(\bar{\theta}_i,\bar{q}_i^\infty)$) and then with P_γ (which, recall, is an element of $\mathcal{P}(Z^\infty)$ for ***any*** $\gamma=(\theta_0,f)$). This results in the following joint distribution over $\Omega \equiv Qx\Theta_0xFxZ^\infty$:

$$\mu_i(.;q_i) \equiv [[1_i(q_i) \otimes \mu_i'''] \otimes P_i(.)] \otimes P_\gamma. \tag{6.6}$$

The measure $\mu_i(.;q_i)$ is the belief over the state space Ω of the agent i with Savage-Bayesian type q_i that chooses behavior strategies according to the rule $\mu_i'''(.;q_i)$. We now record the following:

Lemma 6.7. The mapping from Q_i to $\mathcal{P}(\Omega)$ defined by (6.6) is Borel-measurable; or, equivalently, $\mu_i(.;q_i)$ is a regular conditional probability.

7. "Behind the Veil" or the Ex Ante

7.1. In (6.6) we defined the belief, $\mu_i(.;q_i)$ over Ω of the agent i characterized by the vector q_i. Notice that this probability, although a function of q_i, is not a conditional probability in the usual sense (since we have not defined a joint probability with respect to which the conditioning takes place). On the contrary $\mu_i(.;q_i)$ is a probability on Ω which happens to assign probability one to the given q_i of Q_i. Hence we use the semi-colon, ";" , as opposed to the bar, " | ", when writing $\mu_i(.;q_i)$. At this stage we may consider the primitives of the model to be made up of the economic primitives of section 5, $<I,A,Z^\infty,\Theta,P_\gamma>$ (from which Q_i is constructed) and the behavior strategy choice rules $\{\mu_i'''(. \mid q_i)\}_{i\in I}$. Notice from (6.6) that the measure $\mu_i(.;q_i)$ is ***constructed from, and is uniquely defined by,*** these primitives.

We now go "behind the veil" to an "ex ante" period. We think of this as the period before any agent i "observes" her Savage-Bayesian type q_i. We shall posit the existence of some ex ante belief in this "behind the veil" period. The interpretation will be that at date 0 agents will be "born" and will observe their Savage-Bayesian type, q_i. Agent i's ex post belief, after realization of q_i, will be her ex ante probability conditional upon that realization. The ex post beliefs will be required to respect the fact that q_i uniquely defines agent i's belief over Ω as constructed in (6.6). Formally, we have the following:

7.2. An Ex Ante Subjective Belief, $\mu_i(.)$, for agent i is any $\mu_i \in \mathcal{P}(\Omega)$ such that

$$\mu_i(. \mid q_i) = \mu_i(.;q_i) \quad \text{for } \mu_i\text{-a.e. } q_i, \tag{7.3}$$

where the left-hand-side of (7.3) is (any version of) the probability μ_i conditional on q_i and the right-hand-side is the measure defined in (6.6). A probability $\mu \in \mathcal{P}(\Omega)$ is a ***common prior*** for the agents if for all $i \in I$, μ is an ex ante subjective belief for agent i. By abuse of notation we shall often say the common prior assumption (CPA) holds when we really mean to say that we have a common prior μ for the agents.

7.4. Example.

	$q_B' = (\theta_B', \bar{q}_B^{\infty})$		$q_B'' = (\theta_B'', \bar{q}_B^{\infty})$	
$q_A^* = (\theta_A^*, q_A^{*\infty})$	1/3,	0.4	2/3,	0.4
$q_A^{**} = (\theta_A^{**}, q_A^{**\infty})$	1/4,	0.6	3/4,	0.6

There are two agents A and B. Agent A (resp. B) can have utility parameter θ_A^* or θ_A^{**} (resp. θ_B' or θ_B''). Agent A's belief about B depends upon A's realized utility parameter: When A has utility parameter θ_A^* (resp. θ_A^{**}), A assigns probabilities 1/3 and 2/3 (resp. 1/4 and 3/4) to the event that

B's utility parameter is θ_B' or θ_B''. Regardless of B's utility parameter, B has the same belief about A: Agent B assigns probability 0.4 (resp. 0.6) to agent A having utility parameter θ_A^* (resp. θ_A^{**}). Suppose that the belief of each agent given her utility parameter is common knowledge. Suppose also that there are no parameters of nature and no strategies to choose (or alternatively that there is only one possible value of these items and that value is "common knowledge"). The only relevant orders of beliefs are therefore the first order beliefs of A and B about each other's utility parameters. (All higher order beliefs will be obtained from the common knowledge assumptions we just made.) Let $q_A^{\infty *}$ and $q_A^{\infty **}$ be agent A's belief hierarchy when A has utility parameter θ_A^* and θ_A^{**}, respectively; hence A has two possible Savage-Bayesian types $q_A^* \equiv (\theta_A^*, q_A^{\infty **})$ and $q_A^{**} \equiv (\theta_A^{**}, q_A^{\infty **})$. Regardless of Agent B's utility parameter, B's belief hierarchy will be the same, and we denote this by $\bar{q}_B^{\infty}$; hence, Agent B has two possible Savage-Bayesian types, $q_B' \equiv (\theta_B', \bar{q}_B^{\infty})$ and $q_B'' \equiv (\theta_B'', \bar{q}_B^{\infty})$. There are therefore only four possible states of the world represented by $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ where $\omega_1 \equiv (q_A^*, q_B')$, $\omega_2 \equiv (q_A^*, q_B'')$, $\omega_3 \equiv (q_A^{**}, q_B')$ and $\omega_4 \equiv (q_A^{**}, q_B'')$.

Let $\lambda_A \epsilon (0,1)$ (respectively $\lambda_B \epsilon (0,1)$) denote the ***ex ante*** probability that agent A's (resp. B's) Savage-Bayesian type is q_A^* (resp. q_B'). Then the ex ante beliefs over the state space, μ_A and μ_B resp., will be given by the matrices below:

A's ex ante belief

	q_B'	q_B''
q_A^*	$\lambda_A/3$	$2\lambda_A/3$
q_A^{**}	$(1-\lambda_A)/4$	$3(1-\lambda_A)/4$

B's ex ante Belief

	q_B'	q_B''
q_A^*	$0.4\lambda_B$	$0.4(1-\lambda_B)$
q_A^{**}	$0.6\lambda_B$	$0.6(1-\lambda_B)$

In particular the probabilities assigned to the states $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ by agents A and B are respectively $\{\lambda_A/3, 2\lambda_A/3, (1-\lambda_A)/4, 3(1-\lambda_A)/4\}$ and $\{0.4\lambda_B, 0.4(1-\lambda_B), 0.6\lambda_B, 0.6(1-\lambda_B)\}$.

7.5. An Ex Ante Belief for i is not necessarily an Ex Ante Belief for $j \neq i$. To see this consider the previous example. Fix any $\lambda_B \epsilon (0,1)$ and let μ_B denote the ex ante belief for agent B in that example. Then the conditional probabilities of μ_B given any realization of agent A's Savage-Bayesian type will be given by $\mu_B(\{q_B'\} \mid q_A^*)=\lambda_B$ and $\mu_B(\{q_B'\} \mid q_A^{**})=\lambda_B$. However agent A conditional on her Savage-Bayesian type assigns the event $\{q_B=q_B'\}$ the probability 1/3 or 1/4 according to as her Savage-Bayesian type q_A^* or q_A^{**} respectively; i.e., using the notation of (6.6), $\mu_A(\{q_B=q_B'\};q_A^*)=1/3$ and $\mu_A(\{q_B=q_B'\};q_A^{**})=1/4$. So for μ_B to be an ex ante subjective belief for agent A we must have $\lambda_B=1/3$ ***and*** $\lambda_B=1/4$, which is impossible. Hence μ_B is not an ex ante subjective belief for agent A.

7.6. Common Priors with given Support Need Not Exist. Consider again example 7.4. If μ is a common prior with support equal to the set Ω of that example then μ must be an ex ante subjective belief for B so must equal the probability μ_B constructed in that example for some $\lambda_B \epsilon (0,1)$. However we showed in (7.5) that any such μ_B is ***not*** an ex ante subjective belief for A. Hence there does not exist a common prior with support equal to the set Ω of example 7.4!

7.7. Ex Ante Subjective Beliefs are Without Loss of Generality. There is a sense in which going to the ex ante is ***without*** loss of generality. Agent i, when making decisions will already have knowledge of her Savage-Bayesian type, q_i. All decisions she makes will be conditional on this information. Whatever is her ex ante belief, this conditional will be the same by assumption, and in particular will be given by (7.3). Further, agent i's belief about $j \neq i$ is specified by q_i; after realizing her Savage-Bayesian type q_i, agent i therefore does not really "care" what ex ante belief agent j is using.

Ex ante subjective beliefs with any given support on the set Q_i of Savage-Bayesian types of agent i will always exist. Indeed, fix ***any*** measure $\bar{\mu}_i(.)$ over Q_i. Let $\mu_i(.;q_i)$ be agent i's ex post belief as in (6.6). Then we may define μ_i to be the measure over Ω with marginal over Q_i equal to $\bar{\mu}_i$ and with conditional given q_i equal to $\mu_i(.;q_i)$. (Formally, define for each $q_i \epsilon Q_i$, $\bar{\mu}_i(.;q_i)$ to be the marginal of the measure $\mu_i(.;q_i)$ of (6.6) over the elements of the state space excluding Q_i; i.e., over $Q_{-i} x \Theta_0 x F x Z^\infty$. From lemma 6.7 this is measurable in q_i; use (2.2) and define $\mu_i(.) \equiv \bar{\mu}_i(.) \otimes \bar{\mu}_i(.;q_i)$. This measure μ_i will have support over Q_i equal to that of $\bar{\mu}_i$.) Hence we see that in choosing an ex ante belief for agent i we, the "modeler," have a large number of choices!

One may ask: if going to the ex ante is without much loss of generality, why do we go to the ex ante? There are two reasons. First, by going to the ex ante we have a way of comparing the beliefs of one agent with those of another. For example, the common prior assumption sets the ex ante beliefs of the agents equal to each other. In the next section we shall pursue a weaker version of the common prior assumption (which we will refer to as condition (GH)). Such conditions give us a language within which to say that some properties are true "almost always" when they may not be true "always." For economists the "almost always" statement may be good enough. The very simple example below makes this (perhaps very trivial point) clear. The second reason we use the ex ante beliefs is that it spares us a lot of potential technical problems relating to "measurability" of certain operations and "versions" of conditional probabilities, as for example in section 9 below.

7.8. Example: "For All" versus "For Almost All". There are two agents A and B. Agent B's utility parameter θ_B is an element of $\Theta_B \equiv [0,1]$ which is chosen uniformly from that set. A coin is tossed independently infinitely many times with the probability of HEADS or TAILS equal to θ_B or 1-θ_B respectively. Agent A's utility parameter, θ_A, is the outcome of the infinite coin toss. Agent A does not observe θ_B. If we suppose that the above facts are common knowledge (and further that there are no types of nature or strategies to worry about) then an agent's Savage-Bayesian type may be identified with that agent's utility parameter. We may then construct a fairly trivial state space (essentially equal to $\Omega \equiv \Theta_A x \Theta_B$), and we will obtain a common ex ante belief over Ω for the agents which we denote by μ.

Suppose we are interested in the following ***"question"*** for some reason: "Can agent A infer B's utility parameter from A's realized utility parameter?" Well, agent A will estimate θ_B by computing the long-run average occurrence of HEADS in her realized utility parameter θ_A. From the strong law of large numbers this long-run average will equal θ_B with μ-probability one. There are of course many (actually uncountably many) sequences of HEADS and TAILS for which this is ***not*** the case. Hence we may answer our "question" in the affirmative with μ-probability one and not for all values of the utility or Savage-Bayesian types of the agents.

8. The Outside Observer and Condition (GH)

8.1. We shall say that the collection of subjective ex ante beliefs of agents, $\{\mu_i\}_{i\epsilon I}$, obeys ***condition (GH)*** if μ_i and μ_j are mutually absolutely continuous $\forall i,j\epsilon I$; i.e., if $\forall i,j\epsilon I$ and $\forall$ measurable $D \subseteq \Omega$, $\mu_i(D) > 0$ 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, $\mu_i(. \mid q_i)$ and $\mu_j(. \mid q_j)$, 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.

As a convenience in stating results and theorems it is often useful to posit the existence of an "outside observer" with an ex ante belief μ^* over Ω. The measure μ^* may be interpreted to be the "true" ex ante distribution of the economy if (i) we interpret its marginal on $Qx\Theta_0$ to be the "true" ex ante distribution of the types of agents and nature (whatever that means); if (ii) μ^* is obtained from knowledge of the behavior strategy choice rules of agents, $\{\mu'''(. \mid q_i)\}_{i\epsilon I}$ and if (iii) μ^* is obtained from knowledge of the measure P_γ of (5.6) which determines the distribution over Z^∞ as a function of $\gamma=(\theta_0,f)\epsilon\Theta_0xF$. Alternatively, this same measure could be the ex ante belief of an outside observer (say the economist) whose ***belief*** about the vector of Savage-Bayesian types is equal to the marginal of μ^* over $Qx\Theta_0$ and who knows the items (ii) and (iii) above. When condition (GH) holds it is natural to insist that μ^* be a measure which is mutually absolutely to each of the ex ante subjective beliefs of agents.

9. Date n Belief hierarchies over a Variable, ξ

9.1. Fix any random variable $\xi=\{\xi_0,\{\xi_i\}_{i\epsilon I}\}$ on Ω taking values in some complete and separable metric space $\Xi=\Xi_0x\Pi_{i\epsilon I}\Xi_i$. We shall think of Ξ_i as the space of those coordinates of Ξ pertaining to agent i. In particular we shall think of agent i as having perfect knowledge of the value of ξ_i at each ω. (If there is no such variable for agent i merely set Ξ_i equal to a singleton, and assume it is common knowledge.) Now, for the given space Ξ and the set of agents I, the ***space*** of hierarchies of beliefs over Ξ can be constructed as in section 3. In particular, if we set Y_i and Y_0 in section 3 equal to Ξ_i and Ξ_0 of this section, we obtain hierarchies of beliefs which was referred to in section 3 as $\{B_i^r\}^\infty_{r=1}$ and B_i. We shall for reasons of exposition denote the ***space*** of date n hierarchies of beliefs of agent i over Ξ as $\{B_{in}^r\}_{r=1}^{\infty}$ and B_{in}; (we should really write this as $\{B_{in}^r(\Xi)\}_{r=1}^{\infty}$ and $B_{in}(\Xi)$ to emphasize the dependence on Ξ)!

Fix any date n. We shall use the terminology "at (ω,n)" to mean "at the beginning of date n in the state of the world $\omega\epsilon\Omega$". We proceed to identify for each agent i the hierarchy of beliefs $b_{in}(\omega) \epsilon B_{in}$ over the random variable ξ at (ω,n). This hierarchy of beliefs will specify agent i's belief about ξ at (ω,n); agent i's belief at (ω,n) about the beliefs of others about ξ at (ω,n); agent i's belief at (ω,n) about the beliefs at (ω,n) about beliefs over ξ at (ω,n); etc. In the example of section 1.2 observe that at each date n agent i chooses an action y_{in} which is a function of that agent's date n belief hierarchy over α, the intercept of the demand curve (which is nature's attribute vector).

Define

$$\mathfrak{F}_{i0}=\sigma(\{q_i\}) \quad \text{and} \quad \mathfrak{F}_{in}=\sigma(\{q_i,z_i^{n-1}\}) \text{ for } n\geq 1, \tag{9.2}$$

the σ-algebras on Ω induced by the indicated variables. At the beginning of any date n=0,1,2,... agent i would have observed the vector of variables defining $\mathfrak{F}_{in}$. In (9.2) we define z_i^0 to be the chosen behavior strategy, f_i, of agent i. Since we allow agent i to randomize over her choice of behavior strategies we allow the agent to observe the outcome of that randomization at the beginning of date 1.

Suppose we wanted to construct belief hierarchies over ξ: how might we proceed? Well, fix any collection of ex ante subjective beliefs $\{\mu_i\}_{i\epsilon I}$. Let $\mu_i(. \mid \mathfrak{F}_{in})$ denote any fixed regular version of the conditional probability of μ_i given $\mathfrak{F}_{in}$ and let $\mu_i(. \mid \mathfrak{F}_{in})(\omega)$ denote its value at ω. Let $b_{in}^1(\omega) \epsilon \mathcal{P}(\Xi_{-i})$ denote the distribution of ξ_{-i} induced by the conditional probability $\mu_i(. \mid \mathfrak{F}_{in})(\omega)$; in particular $b_{in}^1(\omega)$ is defined by

$$b_{in}^1(\omega)(D) \equiv \mu_i(\{\xi_{-i}\epsilon D\} \mid \mathfrak{F}_{in})(\omega) \quad \text{for any (measurable) } D\subseteq\Xi_{-i}. \tag{9.3}$$

Eq. (9.3) defines a random variable $b_{in}:\Omega\rightarrow B_{in}^1$. We may wish to consider $b_{in}^1(\omega)$ to be the first order belief of agent i at (ω,n) over the parameter $\xi\epsilon\Xi$. Let us proceed by induction. Suppose that for all $j\epsilon I$ we have defined the r-th order belief of each agent j, $b_{jn}^r:\Omega\rightarrow \mathcal{P}(B_{-in}^{r-1}x\Xi_{-j})$ at ***each*** $\omega\epsilon\Omega$.

Define $b_{in}^{r+1}(\omega)$ to be the probability distribution over the random variables (b_{-in}^r,ξ_{-i}) induced by the conditional probability $\mu_i(.\mid \mathfrak{F}_{in})(\omega)$; in particular $b_{in}^{r+1}(\omega)$ is defined by setting for each $D\subseteq B_{-in}^r x\Xi_{-i}$,

$$b_{in}^{r+1}(\omega)(D)=\mu_i(D'\mid \mathfrak{F}_{in})(\omega) \quad \text{where } D'=\{\omega'\epsilon\Omega:(b_{-in}^r(\omega'),\xi_{-i}(\omega'))\epsilon D\}. \qquad (9.4)$$

By induction we obtain at a sequence $b_{in}(\omega)\equiv\{b_{in}^r(\omega)\}^{\infty}_{r=1}$, for each $i\epsilon I$ at each $\omega\epsilon\Omega$. We would like to refer to $b_{in}(\omega)$ as agent i's belief hierarchy over ξ at (ω,n).

However there are two problems with the construction in (9.3) and (9.4). First, implicit in its definition is the assumption that the ex ante subjective beliefs $\{\mu_i\}_{i\epsilon I}$ are "common knowledge." For example in (9.4) the r+1 th order belief of agent i assumes knowledge of the random variable b_j^r which was constructed using the ex ante belief μ_j. Since ex ante beliefs are arbitrary there is no reason to impose such an assumption. There is a second problem with the definition. The definitions are determined by the ***versions*** of the conditional probabilities used. The construction therefore also assumes that these versions are "common knowledge." Again, since the versions are arbitrary there is no reason to assume they are common knowledge. We will now define "equivalence classes" of the beliefs over ξ to overcome these problems.

Fix any measure $\mu^* \epsilon\mathcal{P}(\Omega)$. We proceed to construct a μ^*-equivalence class of hierarchies of beliefs over ξ. In particular in all of the following definitions we consider μ^* as fixed: Define $C_i(\mu^*)$ to be the set of ex ante subjective beliefs of agent i which are mutually absolutely continuous with respect to μ^*. Then define

$\mathfrak{Z}_{in}^1 \equiv \{b_{in}^1:\Omega\to B_{in}^1 \mid$ (i) b_{in}^1 is $\mathfrak{F}_{in}$-measurable and (ii) for some fixed regular version $\mu_i(. \mid \mathfrak{F}_{in})$ of some ex ante subjective belief for agent i, $\mu_i \epsilon C_i(\mu^*)$, $b_{in}^1(\omega)$ is as defined in (9.3) for μ^*-a.e. $\omega\}$.

Any $b_{in}^1 \epsilon \mathfrak{Z}_{in}^1$ will be referred to as a regular version of the date n first order belief over ξ of agent i. We proceed inductively. Suppose that for some $r \geq 1$ we have defined the μ^*-equivalence class $\mathfrak{Z}_{jn}^r$ of r-th order beliefs for each $j\epsilon I$. Let $\mathfrak{Z}^r_{-in} \equiv \Pi_{j\neq i} \mathfrak{Z}_{jn}^r$. Then we define

$\mathfrak{Z}_{in}^{r+1} \equiv \{b_{in}^{r+1}:\Omega\to B_{in}^{r+1} \mid$ (i) b_{in}^{r+1} is $\mathfrak{F}_{in}$-measurable and (ii) for some fixed regular version $\mu_i(. \mid \mathfrak{F}_{in})$ of some ex ante subjective belief for agent i, $\mu_i \epsilon C_i(\mu^*)$, and for some $b_{-in}^r \epsilon \Pi_{j\neq i} \mathfrak{Z}_{jn}$, $b_{in}^{r+1}(\omega)$ is as defined in (9.4) for μ^*-a.e. $\omega\}$.

Finally, we define $\mathfrak{Z}_{in} \equiv \Pi^{\infty}_{r=1} \mathfrak{Z}_{in}^r$, the μ^*-equivalence class of date n hierarchies of beliefs over ξ of agent i. If we assume that the measure μ^* is mutually absolutely continuous with respect to ***some*** ex ante beliefs $\{\mu_i^*\}_{i\epsilon I}$ of agents, then by using these measures (and any fixed versions of the conditionals given $\mathfrak{F}_{in}$) in (9.3) and (9.4) we see that the equivalence classes $\{\mathfrak{Z}_{in}^r\}^{\infty}_{r=1}$ will all be non-empty in that case. We now have:

Lemma 9.5. (Uniqueness). If b_{in} and $\hat{b}_{in}$ are any two elements of $\mathfrak{Z}_{in}$ then $b_{in} = \hat{b}_{in}$ μ^*-a.e.

Lemma 9.6. (Coherence) Fix any $b_{in} \epsilon \mathfrak{Z}_{in}$. Then for μ^*-a.e. ω, $b_{in}(\omega) \in B_{in}$, where, recall, B_{in} is the set of belief hierarchies over Ξ for agent i which obey the probabilistic coherence property of (3.6).

9.7. Another kind of Coherence. There is another sense in which "coherence" may be desired. Fix any $b_{in} \epsilon \mathfrak{Z}_{in}$. There are two ways to compute agent i's belief at ω about the belief hierarchies of other agents. First, agent i's belief hierarchy at ω, $b_{in}(\omega)$, induces a probability distribution over the belief hierarchies of others via the measure $P_i(b_{in}(\omega))$ of section 4. Alternatively, since agent i knows the (equivalence class of) belief hierarchies of other agents, $b_{-in}(\omega)$, as a function of the state ω, agent i's belief over the belief hierarchies of others may be obtained by "integrating" $b_{-in}(\omega)$ over ω with respect to i's ex ante subjective belief $\mu_i(. \mid \mathfrak{F}_{in})$ conditional on $\mathfrak{F}_{in}$: in particular, the probability agent i assigns to the belief hierarchies of others lying in some set D is $\mu_i(\{\omega' \epsilon \Omega \text{ such that } b_{-in}(\omega') \epsilon D\} \mid \mathfrak{F}_{in})$. (From lemma 9.5 this latter operation is with μ^*-probability one, independent of the ex ante probability μ_i and the version of $b_{-in} \epsilon \mathfrak{Z}_{-in}$.) We may require that these two methods of obtaining agent i's beliefs about the belief hierarchies of other be the same. Well the two methods are indeed the same! From the definition in (4.3), the marginal of $P_i(b_{in}(\omega))$ on $B_{-in}^r x \Xi_{-i}$ is $b_{in}^{r+1}(\omega)$. By the definition in (9.4) this is the same as the distribution induced by $\mu_i(. \mid \mathfrak{F}_{in})$ over the random variable $(b_{-in}^r(\omega), \xi_{-i}(\omega))$. Since probability measures are determined by their values on their (finite-dimensional) coordinates, we obtain the required equivalence.

9.8. The Role of the Ex Ante Subjective Beliefs and Condition (GH). The critical role of the ex ante beliefs and condition (GH) in the above constructions should be obvious. Without the ex ante subjective beliefs it is difficult to even state condition (GH). Attempting to define conditional probabilities ***without*** the ex ante subjective beliefs (e.g., via conditioning with respect to the measure $\mu_i(.;q_i)$ of (6.5)) would result in serious "measurability" headaches.

10. Concluding Remarks

I have provided a framework for studying multi-agent models where there is imperfect information over both the characteristics and the strategies of other agents. In particular, this framework can handle the example of section 1.2. This framework provides a formal mathematical formulation for the original model of incomplete information of Harsanyi (1968).

In the formulation of this paper, modelling the decision-making of agents is stressed. The environment is very "Savage-Bayesian." Hence in many ways this paper is a formalization of Aumann (1987). The work of this paper is continued in a sequel. In Nyarko (1993b) the framework of this paper is used. Various notions of a "type" are introduced there, which are then used to discuss the concept of a Bayesian Nash equilibrium . The sequel also discusses the issues raised in Aumann (1987) on correlated equilibria, using the framework of this paper.

11. Appendix A: Regular Conditional Probabilities

11.1. Let X and W be complete and separable metric spaces. Let $G:X \rightarrow \mathcal{P}(W)$ be any function mapping X into the set of probability measures on W. Let $G(.;x)\epsilon\mathcal{P}(W)$ denote the value of G at x. We shall say that G is a regular conditional probability if for all measurable subsets $D \subseteq W$, $G(D;x)$ is a measurable (real-valued) function of x. (We refer to this as a ***conditional*** probability because we may consider $G(.;x)\epsilon\mathcal{P}(W)$ to be the probability distribution on W "given" or "conditional" on x.) Notice that the measurability in the definition of regularity here is that of the real-valued function $G(D;x)$ which maps X into [0,1]. We often will seek the measurability of the ***function*** $G:X \rightarrow \mathcal{P}(W)$; this by definition will require that for each Borel subset C of $\mathcal{P}(W)$, $G^{-1}(C) \equiv \{x\epsilon X:G(.;x)\epsilon C\}$ is a Borel-measurable subset of X. We now show the following:

Lemma 11.2. Let X, W, G be as above. G is a regular conditional probability if and only if the ***function*** $G:X \to \mathcal{P}(W)$ is Borel-measurable.

Proof of Lemma 11.2: Let D be any closed subset of W. Fix any υ^* in $\mathcal{P}(W)$ and fix any real number $\epsilon > 0$. Define $N(D,\epsilon)(\upsilon^*) \equiv \{\upsilon \text{ in } \mathcal{P}(W): \upsilon(D) < \upsilon^*(D) + \epsilon\}$. A neighborhood for υ^* in the weak topology is a set of the form $\cap^K_{k=1} N(D_k,\epsilon_k)(\upsilon^*)$ where $\{D_k\}^K_{k=1}$ is a ***finite*** collection of closed subsets of W and $\epsilon_k > 0$ for all $k=1,\ldots,K$. (See e.g., Billingsley (1968, p. 236).) Since $\mathcal{P}(W)$ is separable, the Borel subsets of $\mathcal{P}(W)$ are generated by sets of the form $\cap^K_{k=1} N(D_k,\epsilon_k)(\upsilon^*)$ but for a ***countable*** number of υ^* in $\mathcal{P}(W)$. It should be clear that if $G^{-1}(C_m)$ is a measurable subset of X for a countable class of sets $C_m \subseteq \mathcal{P}(W)$, then $G^{-1}(\cap_m C_m)$ is also a measurable subset of X. Hence for $G:X \to \mathcal{P}(W)$ to be measurable it sufficient (and obviously also necessary) that $G^{-1}(C)$ is a Borel-measurable subset of X for each subset $C \subseteq \mathcal{P}(W)$ of the form $N(D,\epsilon)(\upsilon^*)$ defined above for ***fixed*** D, ϵ and υ^*. However, $G^{-1}(N(D,\epsilon)(\upsilon^*)) = \{x \text{ in } X: G(D;x) < \lambda \text{ where } \lambda = \upsilon^*(D) + \epsilon\}$. For any number λ, the regularity of G is a necessary and sufficient condition for the set $\{x \text{ in } X: G(D;x) < \lambda\}$ to be measurable in X. Hence a necessary and sufficient condition for $G:X \to \mathcal{P}(W)$ to be measurable is that G be regular. ▮

11.3. Remark. In the above it should be clear that we could instead endow X with any sub-σ-algebra, $\mathfrak{F}'$, of its Borel σ-algebra. In particular, by trivial modifications of the above proof we may show the following for such $\mathfrak{F}'$: G is $\mathfrak{F}'$-regular (i.e., $G(D;x)$ is $\mathfrak{F}'$-measurable in x for all measurable $D \subseteq W$) if and only if $G:X \to \mathcal{P}(W)$ is $\mathfrak{F}'$-measurable.

11.4. Remark. Let $G:X \to \mathcal{P}(W)$ be as above, and in particular suppose it is Borel-measurable.

Let h:W→R be any uniformly bounded measurable real-valued function. Then from lemma 11.2 it should be easy to verify that by approximating h by simple functions the integral, $\int h(w)G(dw;x)$, of h with respect to G(.;x) is Borel-measurable in x.

11.5. Fix any collection of metric spaces $\{W_k\}^K_{k=1}$ and let $W \equiv \Pi^K_{k=1} W_k$ be the Cartesian product, with $K \leq \infty$. Let $\mathfrak{C}$ be any class of subsets of W. The class $\mathfrak{C}$ is said to be a d-system (or Dynkin system) if

(I) $\mathfrak{C}$ is closed under complementation (i.e., $D \in \mathfrak{C}$ implies $W - D \in \mathfrak{C}$);

(II) $D_1, D_2 \in \mathfrak{C}$ and $D_1 \subseteq D_2$ implies that $D_2 - D_1 \in \mathfrak{C}$; and

(III) $\{D^m\}^\infty_{m=1} \subseteq \mathfrak{C}$ and $D^m \subseteq D^{m+1}$ for all m implies that $\bigcup^\infty_{m=1} D^m \in \mathfrak{C}$.

Given ***any*** class $\mathfrak{C}$ of subsets of W, let $\sigma(\mathfrak{C})$ denote the smallest σ-algebra that contains $\mathfrak{C}$, and let $d(\mathfrak{C})$ denote the smallest d-system that contains $\mathfrak{C}$. Define

$$\mathfrak{R} \equiv \{S \subseteq W: S = \Pi^K_{k=1} S_k \text{ for some Borel-measurable subsets } S_k \subseteq W_k \ \forall k\}. \tag{11.6}$$

$\mathfrak{R}$ is the class of ***measurable rectangles*** of the product space W. The Borel σ-algebra on W is equal to $\sigma(\mathfrak{R})$. We now have the following:

Lemma 11.7. Let $\mathscr{L}$ be a class of Borel subsets of W. Suppose also that $\mathscr{L}$ is a d-system that contains all the measurable rectangles (i.e., $\mathfrak{R} \subseteq \mathscr{L}$). Then $\mathscr{L}$ is equal to the class of Borel subsets on W.

Proof of Lemma 11.7. By assumption $\mathscr{L}$ a class of Borel subsets so $\mathscr{L} \subseteq \sigma(\mathfrak{R})$. A π-system of subsets of W is one which is closed under the formation of finite intersections. It should be clear that $\mathfrak{R}$ is a π-system. It is well-known that if $\mathfrak{Z}$ is a π-system then $\sigma(\mathfrak{Z}) = d(\mathfrak{Z})$; (this is sometimes called the π-λ theorem - see e.g., Cohn (1980), Theorem 1.6.1). Hence $\sigma(\mathfrak{R})=d(\mathfrak{R})$. By assumption $\mathfrak{R}\subseteq\mathcal{L}$, so $d(\mathfrak{R})\subseteq d(\mathcal{L})$. Further, since $\mathcal{L}$ is by assumption a d-system, $d(\mathcal{L})=\mathcal{L}$. In summary we have shown that $\mathcal{L}\subseteq\sigma(\mathfrak{R})=d(\mathfrak{R})\subseteq d(\mathcal{L})=\mathcal{L}$. This implies that $\mathcal{L}=\sigma(\mathfrak{R})$, so proves the claim. ▌

11.8. Let X, W_1 and W_2 be any complete and separable metric spaces. Let $G:X\to\mathcal{P}(W_1)$, $G':W_1\to\mathcal{P}(W_2)$ be any two Borel-measurable mappings on the indicated spaces. Let $G(.;x)$ and $G'(.;w_1)$ denote the values taken by G and G′ at any $x\epsilon X$ and $w_1\epsilon W_1$ respectively. Define $G'':X\to\mathcal{P}(W_1xW_2)$ as follows: at each $x\epsilon X$ the value of G″ at x is $G''(.;x)\equiv G'(.;x)\otimes G(.;w_1)$, the probability measure over W_1xW_2 with a marginal over W_1 equal to $G'(.;x)$ and with a distribution over W_2 conditional on any $w_1\epsilon W_1$ equal to $G'(.;w_1)$. (Lemma 2.3 discusses this $\otimes$ operation and uses 11.7 above to show that this is indeed well-defined.) We have the following:

Lemma 11.9. $G'':X\to\mathcal{P}(W_1xW_2)$ is Borel-measurable.

Proof of lemma 11.9: From Lemma 11.2 it suffices to show that for each measurable subset $D\subseteq W_1xW_2$, $G''(D;x)$ is a Borel-measurable function of x. Define $\mathcal{L}\equiv$ {measurable $D\subseteq W_1xW_2$ | $G''(D;x)$ is a Borel-measurable function of x}. Suppose that $D=D_1xD_2$ for some measurable subsets $D_1\subseteq W_1$ and $D_2\subseteq W_2$. Then

$$G''(D;x)=\int_{\{w_1\epsilon D_1\}} G'(D_2;w_1)dG(.;x) \tag{11.10}$$

where the integration above is over W_1 with respect to the measure $G(.;x)$. Now, $G'(D_2;w_1)$ is by assumption measurable in w_1. Hence from remark 11.4 the integral in (11.10) will be measurable in x. Hence we see that $\mathcal{L}$ contains all measurable rectangles of the form $D=D_1xD_2$. From lemma

11.7 it therefore remains only to show that $\mathcal{L}$ is d-system. However this is easily verified. ▮

12. Appendix B: The Proofs

Proof of Lemma 2.3: Let W, X, G, S and S_x be as in Lemma 2.3. We begin with:

Claim 2.3.1. **(a)** For all $x \epsilon X$, S_x is a measurable subset of W; and **(b)** the function $h: X \rightarrow [0,1]$ defined by $h(x)=G_x(S_x)$ is Borel-measurable.

Proof of Claim 2.3.1: **(a)** Fix any $x \epsilon X$ and define $v: W \rightarrow XxW$ by $v(w)=(x,w)$. The function v is clearly continuous hence is measurable. Since $S_x = v^{-1}(S)$, S_x is therefore measurable subset of W. **(b)** Define $\mathcal{L} \equiv \{$measurable $S \subseteq XxW$: $G_x(S_x)$ is a measurable function of x$\}$. It is very easy to check that $\mathcal{L}$ is a d-system. Hence from lemma 11.7 it suffices to show that $\mathcal{L}$ contains all measurable rectangles in XxW. Fix such a measurable rectangle $S=AxB$, with $A \subseteq X$ and $B \subseteq W$. Let $1_{\{x \epsilon A\}}$ denote the indicator function which is equal to one if $x \epsilon A$ and equal to 0 otherwise. Then $G_x(S_x)=1_{\{x \epsilon A\}}.G_x(B)$. Both $1_{\{x \epsilon A\}}$ and $G_x(B)$ are measurable real-valued functions of x. Hence so is $G_x(S_x)$. Hence $S \epsilon \mathcal{L}$. ▮

Proof of Lemma 2.3 (Cont'd): Part (a) of claim 2.3.1 above implies that for each $x \epsilon X$, $G_x(S_x)$ is well-defined. Part (b) of the claim implies that the integral, $\int G_x(S_x) d\Psi'$ is well-defined. It is easy to check that $\Psi(S) \equiv \int G_x(S_x) d\Psi'$ is a probability measure over (Borel-measurable) subsets S of XxW. This proves part (i) of the lemma. Parts (ii) and (iii) of the lemma follow almost immediately from the definition $\Psi(S) \equiv \int G_x(S_x) d\Psi'$. ▮

Proof of Lemma 3.11: Let r, ρ and D be as in the lemma. Under the probabilistic coherence property, we may re-write the conclusion of the lemma as

$$\Phi_i^{r,r+\rho}(b_i^{r+\rho})(D)=b_i^{r+\rho}(D^{+\rho}). \tag{3.11'}$$

We will prove (3.11') by induction on ρ. When $\rho=1$, $\Phi_i^{r,r+\rho}=\phi_i^r$; (3.11') is then the definition ϕ_i^r. Next suppose that (3.11') holds for some $\rho=1,2,\ldots$. Then from the definition of $\Phi_i^{r,r+\rho+1}$ we have $\Phi_i^{r,r+\rho+1}(b_i^{r+\rho+1})(D)=\Phi_i^{r,r+\rho}(\phi_i^{r+\rho}(b_i^{r+\rho+1}))(D)$; this latter term is, from the probabilistic coherence property, equal to $\Phi_i^{r,r+\rho}(b_i^{r+\rho})(D)$; from the induction hypothesis (i.e., (3.11') for ρ), this is equal to $b_i^{r+\rho}(D^{+\rho})$; from the probabilistic coherence property this is, in turn, equal to $\phi_i^{r+\rho}(b_i^{r+\rho+1})(D^{+\rho})$; which, from the definition of $\phi_i^{r+\rho}$, can be shown to be equal to $b_i^{r+\rho+1}(D^{+(\rho+1)})$. Combining all our arguments results in $\Phi_i^{r,r+\rho+1}(b_i^{r+\rho+1})(D)=b_i^{r+\rho+1}(D^{+(\rho+1)})$. This is (3.11') for $\rho+1$. Hence (3.11') is true for all $\rho=1,2,\ldots$ ∎

Proof of Lemma 6.7.: This follows immediately from repeated use of lemma 11.9. Indeed, suppose that in lemma 11.9 we set $X=Q_i$, $W_1=Q_i$, $W_2=F_i$, and also set G(.;x) equal to the measure which assigns probability one to the indicated x and set $G'(.;w_1)$ equal to $\mu_i'''(.;q_i)$ for $w_1=q_i$. Then we may conclude that the measure $\mu_i^1(.;q_i)\equiv[1_i(q_i)\otimes\mu_i''']$ over $Q_i x F_i$ is Borel-measurable in q_i. Next, set $X=Q_i$, $W_1=Q_i x F_i$, $W_2=Q_{-i}x\Theta_0 x F_{-i}$, $G(.;x)=\mu_i^1(.;q_i)$ for $x=q_i$ and $G'(.;w_1)$ equal to $P_i(\hat{q}_i^\infty)$ for any $w_1=((\hat{\theta}_i,\hat{q}_i^\infty),\hat{f}_i)$. Then we may conclude that the measure $\mu_i^2(.;q_i)\equiv\mu_i^1(.;q_i)\otimes P_i(.)$ $=[[1_i(q_i)\otimes\mu_i''']\otimes P_i(.)]$ over $Qx\Theta_0 xF$ is Borel-measurable in q_i. One more use of lemma 11.9 (with G' defined via P_γ) proves the lemma. ∎

Proof of Lemma 9.5. It suffices to show that for each r and for any b_{in}^r and $\hat{b}_{in}^r$ in $\mathfrak{Z}_{in}^r$, $b_{in}^r=\hat{b}_{in}^r$ μ^*-a.e. This can be proved by induction on r from the definitions of the sets $\mathfrak{Z}_{in}^r$ and the claim below:

Claim 9.5.1. Let μ_i and $\hat{\mu}_i$ be any two ex ante subjective beliefs of agent i and suppose that they are mutually absolutely continuous with respect to some measure $\mu^*\epsilon\mathcal{P}(\Omega)$. Let $\mu_i(.\mid\mathfrak{F}_{in})$ and $\hat{\mu}_i(.\mid\mathfrak{F}_{in})$ denote any fixed regular versions of the conditionals of μ_i and $\hat{\mu}_i$ given $\mathfrak{F}_{in}$. Let x and $\hat{x}$ be any two random variables on Ω taking values in some complete and separable metric space X, and suppose that $x=\hat{x}$ μ^*-a.e. Then **(i)** $\mu_i(.\mid\mathfrak{F}_{in})=\hat{\mu}_i(.\mid\mathfrak{F}_{in})$ μ^*-a.e., and **(ii)** the distribution over x induced by $\mu_i(.\mid\mathfrak{F}_{in})(\omega)$ is equal to the distribution over $\hat{x}$ induced by $\hat{\mu}_i(.\mid\mathfrak{F}_{in})(\omega)$ for μ^*-a.e. ω; (i.e., for μ^*-a.e. ω, $\mu_i(\{x\epsilon D\}\mid\mathfrak{F}_{in})(\omega)=\hat{\mu}_i(\{\hat{x}\epsilon D\}\mid\mathfrak{F}_{in})(\omega)$ for all measurable $D\subseteq X$).

Proof of Claim 9.5.1. **(i)** From the definition (see (7.2)) of an ex ante subjective belief the conditionals of μ_i and $\hat{\mu}_i$ given q_i will be equal, μ^*-a.e. It is easily verified that this in turn implies that $\mu_i(.\mid\{q_i,z_i^{n-1}\})=\mu_i(.\mid\{q_i,z_i^{n-1}\})$ μ^*-a.e. Since $\mathfrak{F}_{in}=\sigma(\{q_i,z_i^{n-1}\})$, this proves part (i) the claim. **(ii)** This follows immediately from part (i). ∎

Proof of Lemma 9.6: Fix any $i\epsilon I$. Let $b_{in}=\{b_{in}^r\}_{r=1}^{\infty}\epsilon\mathfrak{Z}_{in}$. We seek to show that for all $r\geq 1$ and for all $b_{in}^{r+1}\epsilon\mathfrak{Z}_{in}^{r+1}$ and $b_{in}^r\epsilon\mathfrak{Z}_{in}^r$, $\phi_i^r(b_{in}^{r+1}(\omega))=b_{in}^r(\omega)$ for μ^*-a.e. ω. We shall prove this by induction on r. Fix any $b_{in}^2\epsilon\mathfrak{Z}_{in}^2$ and $b_{in}^1\epsilon\mathfrak{Z}_{in}^1$. From the definition of the equivalence class $\mathfrak{Z}_{in}^2$, $b_{in}^2(\omega)$ is the distribution of the random variables (b^1_{-in},ξ_{-i}) for some $b^1_{-in}\epsilon\mathfrak{Z}^1_{-in}$ induced by some version $\hat{\mu}_i(.\mid\mathfrak{F}_{in})(\omega)$ of the conditional probability of some $\hat{\mu}_i\epsilon C_i(\mu^*)$. Hence from the definition of ϕ_i^1 in (3.4), $\phi_i^1(b_{in}^2(\omega))$ is the distribution of ξ_{-i} induced by $\hat{\mu}_i(.\mid\mathfrak{F}_{in})(\omega)$. By definition of $\mathfrak{Z}_{in}^1$, $b_{in}^1(\omega)$ is also a distribution of ξ_{-i}, but is the distribution induced by some version of the conditional probability, $\bar{\mu}_i(.\mid\mathfrak{F}_{in})$, of some possibly different $\bar{\mu}_i\epsilon C_i(\mu^*)$. Claim 9.5.1(i) therefore implies that

$\phi_i^1(b_{in}^2(\omega))=b_{in}^1(\omega)$ μ^*-a.e. Next, suppose that we have shown for some $r\geq 2$ that $\forall j\epsilon I$, $\forall b_{jn}^r \epsilon \mathfrak{Z}_{jn}^r$ and $\forall b_{jn}^{r-1} \epsilon \mathfrak{Z}_{jn}^{r-1}$,

$$\phi_j^{r-1}(b_{jn}^r(\omega))=b_{jn}^{r-1}(\omega) \ \mu^*\text{-a.e.} \tag{9.6.1}$$

By definition of $\mathfrak{Z}_{in}^{r+1}$, $b_{in}^{r+1}(\omega)$ is the distribution of the random variables $(\hat{b}^r_{-in},\xi_{-i})$ for some $\hat{b}^r_{-in}\epsilon\mathfrak{Z}^r_{-in}$ induced by some version $\hat{\mu}_i(. \mid \mathfrak{F}_{in})(\omega)$ of the conditional probability of some $\hat{\mu}_i\epsilon C_i(\mu^*)$. Hence, from the definition of the operator ϕ_i^r, $\phi_i^r(b_{in}^{r+1}(\omega))$ is the distribution of the random variables $(\phi_{-i}^{r-1}(\hat{b}^r_{-in}),\xi_{-i})$ induced by $\hat{\mu}_i(. \mid \mathfrak{F}_{in})(\omega)$. On the other hand, by definition of $\mathfrak{Z}_{in}^r$, $b_{in}^r(\omega)$ is the distribution of the pair of $(\bar{b}_{-in}^{r-1},\xi_{-i})$ for some $\bar{b}_{-in}^{r-1}\epsilon\mathfrak{Z}_{-in}^{r-1}$ induced by some version $\bar{\mu}_i(. \mid \mathfrak{F}_{in})$ of the conditional probability of some $\bar{\mu}_i\epsilon C_i(\mu^*)$. From the induction hypothesis $\phi_{-i}^{r-1}(\hat{b}^r_{-in})=\bar{b}_{-in}^{r-1}$ μ^*-a.e. Hence claim 9.5.1(b) implies that $\phi_i^r(b_{in}^{r+1}(\omega))=b_{in}^r(\omega)$ μ^*-a.e. This is the induction step (9.6.1) for $r+1$. By induction this is true for all $r\geq 1$. ▮

13. References

Ambruster, W. and W. Boge (1979): "Bayesian Game Theory," in "Game Theory and Related Topics," eds. O. Moeschlin and D. Pallachke, pp. 17-28, North Holland, Amsterdam.

Aumann, R. (1987):"Correlated Equilibrium as an Expression of Bayesian Rationality," Econometrica, 55, 1-18.

Billingsley, P. (1968): Convergence of Probability Measures, Wiley, New York.

Boge, W. and Th. Eisele (1979):"On Solutions of Bayesian Games," International Journal of Game Theory, 8(4), pp. 193-215.

Brandenberger, A. and E. Dekel (1993):"Hierarchies of Beliefs and Common Knowledge," Journal of Economic Theory, 59(1), 189-198.

Cohn, D. (1980):"Measure Theory," Birkhauser, Boston.

Chow, Y.S. and H. Teicher (1978): "Probability Theory," Springer Verlag, New York.

Feldman, M. (1988): "An Example of Convergence to Rational Expectations with Heterogeneous Beliefs," International Economic Review, 28, 635-650.

Harsanyi, J.C. (1967,1968): "Games with Incomplete Information Played by Bayesian Players," Parts I,II,III, Management Science, vol. 14, 3,5,7.

Heifetz, A. (1990):"The Bayesian Formulation of incomplete information - The noncompact case," School of Mathematical Sciences, Tel Aviv University.

Jordan, J. S. (1991a): "Bayesian Learning in Normal Form Games," Games and Economic Behavior, 3, 60-81.

------------- (1991b): "Bayesian Learning In Repeated Games," Manuscript, University of Minnesota.

Kalai, E. and E. Lehrer (1990):"Bayesian Learning and Nash Equilibrium," Manuscript, Northwestern University.

Mertens, J.-F., and S. Zamir (1985):"Formalization of Bayesian Analysis for Games with incomplete Information," International Journal of Game Theory, 14:1-29.

Myerson, R. (1985):"Bayesian Equilibrium and Incentive-Compatibility: An Introduction," in "Social Goals and Social Organization, ed. by L. Hurwicz, D. Schmeidler, and H. Sonnenchein. New York: Cambridge University Press.

Nyarko, Y. (1991a):"Most Games Violate the Harsanyi Doctrine," C.V. Starr Working Paper No. 91-39, New York University.

-------- (1991b):"The Convergence of Bayesian Belief Hierarchies," C.V. Starr Center Working Paper No. 91-50, New York University.

-------- (1992):"Bayesian Learning Without Common Priors and Convergence to Nash equilibrium," C.V. Starr Working Paper No. 92-25, New York University.

--------- (1993a):"Bayesian Learning in Leads to Correlated Equilibria in Normal Form games," Economic Theory (forthcoming).

--------- (1993b):"The 'Types' of a Bayesian Equilibrium" Manuscript, New York University.

Parthasarathy, K. P. (1967):"Probability Measures on Metric Spaces," Academic Press, New York.

Savage, L.J (1954): The Foundations of Statistics, New York, Wiley.

Tan, T. and S. Werlang (1988):"The Bayesian Foundations of Solution Concepts of Games," Journal of Economic Theory, 45, 370-391.

--------------------- (1988):"On Aumann's Notion of Common Knowledge -- An alternate approach,"CARESS Working Paper No. 88-09, University of Pennsylvannia.

Townsend, R. (1978):"Market Anticipations, Rational Expectations, and Bayesian Analysis, "International Economic Review, 19(2), 481-494.