Discussion Paper #710

WE EVENTUALLY AGREE

by

James Bergin*

September 1986

*Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60201.

[8]

WE EVENTUALLY AGREE

This paper is concerned with convergence of beliefs under iterative processes of information exchange. Given different agents with different information, those agents will typically attach different probabilities to any given event. Geanakoplos and Polemarchakis (1982) showed that, with the information of each player represented as a partition of states of the world, if each player's information partition is finite and if agents iteratively announce the value of their posterior distributions then the value of each agent's posterior distribution converges to a common value. Nielsen (1984) extended this result to a more general information structure. With the information of each agent represented as a Boolean sigma algebra, a more general communication process and possibly an infinite number of agents, Nielsen showed that the posteriors of agents converge almost surely to a common posterior. Recently, McKelvey and Page (1986) generalized the result of Geanakoplos and Polemarchakis to an iterative process of public announcement: at each stage individuals compute their posteriors, a (scalar) statistic based on these posteriors is computed and announced. Then, at the next stage, individuals compute new posteriors, conditional on the value of this statistic and their information. They show that with this iterative process and finite information partitions, convergence to a common posterior occurs.

The purpose of this paper is to extend the results of Geanakoplos and Polemarchakis and McKelvey and Page to the case where agents' information structures are represented by sigma algebras. In many economic models, information of agents is represented as the observation of a signal or random variable. In such circumstances, the conditional probability of an event, given the random variable, is understood the mean the conditional probability of the event, given the sigma algebra generated by the random variable. In this case it is natural to represent the informational content of a random variable by the sigma algebra it determines. Similarly, the conditional probability of an event, given a partition is understood to mean the conditional probability of the event, given the sigma algebra generated by the partition. To see that partitions may not adequately represent information when conditional probabilities or expectations must be defined, consider the following example.

Let $\Omega = [0,1]$, let $\mathcal{F}$ be the Borel field and let p be lebesgue measure on $[0,1]$. Define the random variable x as $x(\omega) = \omega$, $\omega \in [0,1]$. Thus the partition of $[0,1]$ determined by x is the set of all points in $[0,1]$. This random variable or the partition determined by it correspond to what one means by full information: knowing that the random variable has the value $\bar{x}$, one can infer that the value of ω is $\bar{\omega} = x^{-1}(\bar{x})$ and knowing the element of the partition that contains ω is an equivalent to knowing the value of ω. Let $\mathcal{F}_x$ be the sigma algebra generated by x ($\mathcal{F}_x = \mathcal{F}$ here) and $\mathcal{F}_p$ the sigma algebra generated by the partition. Also, let $\mathcal{F}_n = \{\{\emptyset\}, \Omega\}$, $\mathcal{F}_n$ corresponds to no information.

Given any $A \in \mathcal{F}$,

$$p(A|\mathcal{F}_p) = p(A) = p(A|\mathcal{F}_n), \text{ a.s.}$$

while

$$p(A|\mathcal{F}_x) = 1_A, \text{ a.s.}$$

where 1_A is the indicator function of A.

Also, given any integrable random variable y,

$$E(y|\mathcal{F}_p) = E(y) = E(y|\mathcal{F}_n), \text{ a.s.}$$

while

$$E(y|\mathcal{F}_x) = y, \text{ a.s.}$$

Yet another way of viewing the issue is to consider maximizing a state dependent utility function conditional on the different sigma algebras. Let $u(x,\omega) = 1 - (x - \omega)^2$. Then

$$\text{MaxE}\{u(x,\omega)|\mathcal{F}_p\} \text{ gives } x(\omega) = 1/2, \text{ a.s.}$$

and

$$\text{MaxE}\{u(x,\omega)|\mathcal{F}_x\} \text{ gives } x(\omega) = \omega, \text{ a.s.}$$

The decision function $x(\omega) = \omega$ a.s. (and not $x(\omega) = 1/2$ a.e.) corresponds naturally to the decision rule of an informed individual. However, any function which is $\mathcal{F}_p$ measurable is constant almost surely.

It is clear that the fully informative partition does not determine a conditional distribution which reflects full information, whereas the algebra $\mathcal{F}_x$ does. (Note, however, that the algebra $\mathcal{F}_x$ is not generated by any partition, since it contains all the singleton sets.) Furthermore, since any decision function which is $\mathcal{F}_p$ measurable is constant almost surely, this conflicts with the intuition that one wishes to allow a decision function to vary with information.[1] These observations make clear that in this example $\mathcal{F}_x$ is the appropriate algebra to represent the information and makes clear the need to discuss the convergence of beliefs when information is represented by sigma algebras. In section 2 the framework for the discussion is given. Section 3 provides the extension of the result of Geanakoplos and Polemarchakis. This is essentially the same result as that given by Nielsen although the way in which information is represented is slightly different. The main reason for providing this discussion is that it serves as an introduction to the work of McKelvey and Page which is the main focus of this paper. The extension of the work of McKelvey and Page to the case where information is represented by sigma algebras is taken up in section 4. It should be pointed out that both Nielsen and McKelvey and Page discuss other issues in addition to convergence of beliefs. Here the focus is exclusively on their results with regard to convergence of beliefs.

2. The Model

A probability space $(\Omega,\mathcal{F},p)$ is given where p is the prior distribution over Ω, common to all individuals. The set of agents is denoted N and the information of each agent $i \in N$ is represented by a sub-sigma algebra $\mathcal{F}_i$ of $\mathcal{F}$. Information represented in this way has the following informal interpretation: given $\omega \in \Omega$, for any $F \in \mathcal{F}_i$, agent i "knows" if $\omega \in F$ or $\omega \in F^c$ (F^c is the complement of F). Given an event $A \in \mathcal{F}$ a conditional probability with respect to $\mathcal{F}_i$ is an $\mathcal{F}_i$ measurable, integrable function $p(A|\mathcal{F}_i)$ which satisfies the functional equation $\int_F p(A|\mathcal{F}_i)dp = p(A \cap F)$, for all $F \in \mathcal{F}_i$. Denote by $\hat{\mathcal{F}} = \bigcap_{i\in N} \mathcal{F}_i$ the maximal sigma algebra contained in all of the $\mathcal{F}_i$'s and by $\check{\mathcal{F}} = \bigvee_{i\in N} \mathcal{F}_i$ the minimal sigma algebra containing all the $\mathcal{F}_i$'s. (Thus, $F \in \hat{\mathcal{F}}$ implies that $F \in \mathcal{F}_i$ for all $i \in N$.)

3. Iterative Announcement of Posterior Distributions

The situation considered here is that where agents announce their posteriors on an event A at each stage and iteratively update their posteriors conditional on their private information and the past history of announced posteriors.

The iterative process is described as follows: at stage 1 agent i announces $p(A|\mathcal{F}_i) = q_i^{(1)}$, $i \in N$. The information revealed by these announcements is represented by $\mathcal{A}_1 = \bigvee_{i \in N} \mathcal{A}_{i1}$ where $\mathcal{A}_{i1} = \sigma(q_i^{(1)})$. At stage 2 agent i computes and announces $p(A|\mathcal{F}_i \vee \mathcal{A}_1) = q_i^{(2)}$. Let $\mathcal{A}^t = \bigvee_{s=1}^{t} \mathcal{A}_s$, $\mathcal{A}_s = \bigvee_{i \in N} \mathcal{A}_{is}$ and $a_{is} = \sigma(q_i^{(s)})$. At stage t agent i computes and announced $p(A|\mathcal{F}_i \vee \mathcal{A}^{t-1}) = q_i^{(t)}$.

Theorem 1: $q_i^{(t)}$ converges almost surely, say to $q_i^{(\infty)}$, for all $i \in N$ and $q_i^{(\infty)} = q_j^{(\infty)}$ almost surely, for all $i,j \in N$.

Proof: Let $\mathcal{A}^\infty = \bigvee_{s \geq 1} \mathcal{A}_s$ and let $q_i^{(\infty)} = p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)$. Since $\mathcal{F}_i \vee \mathcal{A}^{t-1} \subset \mathcal{F}_i \vee \mathcal{A}^\infty$,

$$E\{q_i^{(\infty)}|\mathcal{F}_i \vee \mathcal{A}^{t-1}\} = E\{E\{I_A|\mathcal{F}_i \vee \mathcal{A}^\infty\}|\mathcal{F}_i \vee \mathcal{A}^{t-1}\}$$

$$= E\{I_A|\mathcal{F}_i \vee \mathcal{A}^{t-1}\} = p(A|\mathcal{F}_i \vee \mathcal{A}^{t-1})$$

$$= q_i^t \text{ a.s. (almost surely).}$$

Therefore, $(q_i^{(t)}, \mathcal{F}_i \vee \mathcal{A}^{t-1})$ is a bounded martingale so that $q_i^{(t)}$ converges almost surely to $q_i^{(\infty)} = p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)$.

Next, since $\mathcal{A}^t = \bigvee_{s=1}^{t} \mathcal{A}_s$, $\mathcal{A}_s = \bigvee_{i \in N} \mathcal{A}_{is}$ and $\mathcal{A}_{it} = \sigma(q_i^{(t)})$, $q_i^{(t)}$ is measurable at $\mathcal{A}^t$. Therefore,

$$E\{q_i^{(t)} | \mathcal{A}^t\} = q_i^{(t)} \text{ a.s.}$$

Thus,

$$E\{q_i^{(t)} | \mathcal{A}^t\} \to q_i^{(\infty)} \text{ a.s.}$$

and

$$E\{q_i^{(t)} | \mathcal{A}^t\} \to E\{q_i^{(\infty)} | \mathcal{A}^\infty\} \text{ a.s.,}$$

so that

$$E\{q_i^\infty | \mathcal{A}^\infty\} = q_i^\infty \text{ a.s.}$$

and

$$\begin{aligned} E\{q_i^\infty | \mathcal{A}^\infty\} &= E\{E\{I_A | \mathcal{F}_i \vee \mathcal{A}^\infty\} | \mathcal{A}^\infty\} \\ &= E\{I_A | \mathcal{A}^\infty\} \\ &= p(A | \mathcal{A}^\infty) \text{ a.s.} \end{aligned}$$

Thus,

$$q_i^\infty = p(A | \mathcal{A}^\infty) \text{ a.s.,} \quad \forall\, i \in N. \qquad \square$$

This iterative process involves each agent obtaining a substantial amount of information at each stage--the posterior distribution of each agent. McKelvey and Page restrict the amount of information made available to agents. Instead of the posterior distributions being announced at each stage, a (scalar) statistic based on the posterior distributions is announced with all agents having finite information partitions. For a class of admissible statistics an iteration similar to the above leads to convergence, almost everywhere, to a common posterior distribution. The next section is concerned with extending the result to a more general information structure.

4. Iterative Announcement of an Admissible Statistic on Prior Distributions

The discussion of this section follows closely that of McKelvey and

Page. Central to extending their result is the introduction of regular conditional distributions which are proper almost everywhere. Given a measurable space $(\Omega, \mathcal{F})$, a regular conditional probability distribution on $\mathcal{F}$ given $\mathcal{A} \subseteq \mathcal{F}$ which is almost everywhere proper is a function p on $\Omega \times \mathcal{F}$ satisfying the following three properties:

(i) for each $\omega \in \Omega$, $p(\cdot, \omega)$ is a probability measure on $\mathcal{F}$;

(ii) for each $F \in \mathcal{F}$, $p(F, \cdot)$ is an $\mathcal{A}$ measurable function with $p(F, \cdot) = p(F|\mathcal{A})$ a.e.;

(iii) $\exists\ A \in \mathcal{A}$ with $p(A) = 0$ such that $p(G, \omega) = 1$ if $\omega \in G \in \mathcal{A}$, $\omega \notin A$.

To proceed, some notation and definitions are required. These are taken directly from the paper of McKelvey and Page.

Let $I = [0,1]$, $I^n = [0,1]^n$, $M(I)$ and $M(I^n)$ the set of probability measures on I and I^n, respectively.

<u>Definition 1</u>: If $\lambda, \mu \in M(I)$, λ <u>stochastically dominates</u> μ (written $\lambda \geqslant \mu$) iff $\lambda([0,b]) \leqslant \mu([0,b])$, $\forall\ b \in [0,1]$.

Given $\mu \in M(I^n)$, define $\mu_i \in M(I)$ from μ by $\mu_i(C) = u(I \times \ldots C \times \ldots I)$, $\forall\ C \in B(I)$. Here $B(I)$ is the borel sigma algebra on I and C is the i^{th} coordinate of the product $I \times \ldots C \times \ldots I$.

Then λ <u>stochastically dominates</u> μ, $(\lambda, \mu \in M(I^n))$ if $\lambda_i \geqslant \mu_i$, $i = 1, \ldots, n$. This is written $\lambda \geqslant \mu$. If $\lambda \geqslant \mu$ and $\lambda \neq \mu$, this is denoted by $\lambda > \mu$.

<u>Definition 2</u>: $f\colon I^n \to \mathbb{R}$ is <u>stochastically monotone</u> iff $\forall\ \lambda, \mu \in M(I^n)$, $\lambda > \mu \Rightarrow f(\lambda) > f(\mu)$, where $f(\lambda) = \int f d\lambda$ and $f(\mu) = \int f d\mu$.

<u>Definition 3</u>: $h\colon I^n \to \mathbb{R}$ is <u>stochastically regular</u> iff $h = g \circ f$ where f is

stochastically monotone and $g: \mathbb{R} \to \mathbb{R}$ is invertible on the range of g.

Given any function h, $h: I^n \to \mathbb{R}$, the iterative process of public announcement may be described.

Stage 1

Agent i computes $q_i^1 = p(A|\mathcal{F}_i)$

$\phi^1 = h(q^1)$ is publicly announced, $q^1 = (q_1^1,\dots,q_n^1)$.

Let $\mathcal{A}_1 = \sigma(\phi^1)$.

Stage 2

Agent i computes $q_i^2 = p(A|\mathcal{F}_i \vee \mathcal{A}^1)$

$\phi^2 = h(q^2)$ is publicly announced, $q^2 = (q_1^2,\dots,q_n^2)$.

Let $\mathcal{A}_2 = \sigma(\phi^2)$ and $\mathcal{A}^2 = \mathcal{A}_1 \vee \mathcal{A}_2$.

Stage t

Agent i computes $q_i^t = p(A|\mathcal{F}_i \vee \mathcal{A}^{t-1})$

where $\mathcal{A}^{t-1} = \bigvee_{s=1}^{t-1} \mathcal{A}_s$, $\mathcal{A}_s = \sigma(\phi^s)$

$\phi^t = h(q^t)$ is publicly announced.

The interest is in the limiting behavior of $\{q_i^t\}_{i \in N}$.

To proceed, some additional structure will be imposed on the measurable space $(\Omega,\mathcal{F})$, introduced in section 2. It will be assumed that Ω is a complete separable metric space and $\mathcal{F}$ the set of borel subsets of Ω. Then given a stochastic process $(x_1,x_2,\dots,x_t,\dots) = x$, if $\mathcal{A} = \sigma(x)$ there exists are regular conditional distribution on $\mathcal{F}$ given $\mathcal{A}$ which is almost everywhere proper.

The conditions under which such a probability exists are quite general and are discussed in Blackwell (1955), Blackwell and Ryll-Nardzewski (1963) and Blackwell and Dubins (1975). Recently, Brandenburger and Dekel (1985) have used properness of regular conditional distributions in the context of a common knowledge framework.

The theorem may now be stated. Recall the context of the discussion: a probability space $(\Omega, \mathcal{F}, p)$, an information structure $\mathcal{F}_1, \ldots, \mathcal{F}_n$ and an aggregation function $h: I^n \to \mathbb{R}$.

Theorem 2: For all i, q_i^t converges to q_i^∞ (say), almost surely. If h is continuous and satisfies stochastic regularity then $q_i^\infty = q_j^\infty$ almost surely, for all $i,j \in N$.

Proof: An important part of the proof involves extending a theorem of McKelvey and Page from the case where information is represented by partitions to the case where information is represented by sigma algebras. This is done in Proposition 1 below.

For the following proposition it is assumed that there exists a stochastic process $(x_1, x_2, \ldots, x_t, \ldots) = x$ such that $\mathcal{A} = \sigma(x)$.

Proposition 1: Let $\phi = h(q)$ satisfy stochastic regularity. If $\mathcal{A} \subseteq \mathcal{F}$ satisfies $\mathcal{A} \subseteq \bigcap_{i \in N} \mathcal{F}_i$ and $\sigma(\phi) \subseteq \mathcal{A}$, then for almost all $\omega \in \Omega$, $q_i(\omega) = p(A|\mathcal{A})(\omega)$ where $q_i = p(A|\mathcal{F}_i)$.

Proof: If $p(A) = 0$, then $\forall\, i \in N$

$$\int_\Omega p(A|\mathcal{F}_i)dp = 0$$

so that $p(A|\mathcal{F}_i) = 0$ a.e. $\forall\, i \in N$. Similarly,

$$\int_\Omega p(A|\mathcal{A})dp = 0$$

so that $p(A|\mathcal{A}) = 0$ a.s. and the proposition holds.

Suppose that $p(A) > 0$. Define the measure p_A on $\mathcal{F}$ by

$$p_A(F) = p(A \cap F)/p(A), \ \forall\ F \in \mathcal{F}$$

Fix regular conditional probabilities on $\mathcal{F}$ given $\mathcal{A}$ which are almost everywhere proper. This gives two functions $p(F,\omega)$ and $p_A(F,\omega)$ defined for $F \in \mathcal{F}$, $\omega \in \Omega$, where, for example:

i. for each $\omega \in \Omega$, $p(F,\omega)$ is a probability measure on $\mathcal{F}$;

ii. for each $F \in \mathcal{F}$, $p(F,\omega)$ is an $\mathcal{A}$ measurable function with $p(F,\omega) = p(F|\mathcal{A})(\omega)$ a.e.

iii. $\exists\ V \in \mathcal{A}$, $p(V) = 0$ such that

$$p(G,\omega) = 1 \text{ if } \omega \in G \in \mathcal{A},\ \omega \notin V$$

The properties of $p_A(F,\omega)$ are the same except that in (iii) the set V is replaced by a set V_A with $p_A(V_A) = 0$.

Next, $\forall\ \omega \in \Omega$ let

$$\lambda(C,\omega) = p_A(q^{-1}(C),\omega),\ C \in B(I^n)$$

and

$$\mu(C,\omega) = p(q^{-1}(C),\omega)$$

For each $\omega \in \Omega$, define measures $\lambda_i(\cdot,\omega)$, $\mu_i(\cdot,\omega)$ on $B(I)$ by

$$\lambda_i(C,\omega) = \lambda(I \times \ldots C \times \ldots I,\omega),\ C \in B(I)$$

$$\mu_i(C,\omega) = \mu(I \times \ldots C \times \ldots I,\omega),\ C \in B(I)$$

where C occupies the i^{th} position in $I \times \ldots C \times \ldots I$.

The following two lemmas will be used in the proof of Proposition 1, these lemmas are proved in the paper of McKelvey and Page.

Lemma 1: Let $\lambda,\mu \in M(I)$. Suppose that there exists $\phi: \mathbb{R} \to \mathbb{R}$, monotone on I with

$$\lambda(C) = \int_C \phi d\mu,\ \forall\ C \in B(I)$$

Then $\lambda \geqslant \mu$ with $\lambda > \mu$ unless $\mu(\{t\}) = 1$.

Lemma 2: Let $\lambda_i(C,\omega)$, $\mu_i(C,\omega)$ be defined as above. Then

$$p(A|\mathcal{A})(\omega)\lambda_i(C,\omega) = \int_C t\mu_i(dt,\omega) \text{ a.e. } \omega \in \Omega$$

In the following, the expression "p a.e." means that the statement holds for all $\omega \in \Omega$ except on a set of p measure 0. Similarly, the expression "$p(\cdot,\omega^*)$ a.e." means that the statement holds for all $\omega \in \Omega$ except on a set of $p(\cdot,\omega^*)$ measure 0.

Next, it will be shown that p a.e. $\omega^* \in \Omega$

$$\lambda(\cdot,\omega^*) > \mu(\cdot,\omega^*) \text{ unless q is constant } p(\cdot,\omega^*) \text{ a.e.}$$

To see this there are two cases to consider:

Case 1: $p(A|\mathcal{A})(\omega^*) = 0$. If ω^* is in the p null set on which the equality in

Lemma 2 does not hold for $i \in N$, ignore it. Otherwise, by Lemma 2

$$\int_C t\mu_i(dt,\omega^*) = 0, \ \forall \ C \in B(I), \ \forall \ i \in N$$

Therefore, $\mu_i(\{0\},\omega^*) = 1$, $\forall\ i \in N$ (so $\lambda(\cdot,\omega^*) \geqslant \mu(\cdot,\omega^*)$) and $p(q_i^{-1}(0),\omega) = 1$, $\forall\ i \in N$. (Since $\mu_i(\{0\}, \omega^*) = p(q_i^{-1}(0), \omega^*)$.)

Consequently, $\forall\ i \in N$, $q_i = 0$, $p(\cdot,\omega^*)$ a.e. and so $q = 0$, $p(\cdot,\omega^*)$ a.e.

<u>Case 2</u>: $p(A|\mathcal{A})(\omega^*) > 0$. Again, ignore the p null set on which the equality in Lemma 2 does not hold for all $i \in N$. Since $p(A|\mathcal{A})(\omega^*) > 0$,

$$\lambda_i(C,\omega^*) = \int_C \{\frac{t}{p(A|\mathcal{A})(\omega^*)}\} \ \mu_i(dt,\omega^*), \ \forall \ C \in B(I), \ \forall \ i \in N$$

Applying Lemma 1 gives, $\forall\ i \in N$ there exists t_i such that

$$\lambda_i(\cdot,\omega^*) \geqslant \mu_i(\cdot,\omega^*) \text{ with } \lambda_i(\cdot,\omega^*) > \mu_i(\cdot,\omega^*)$$

unless $\mu_i(\{t_i\},\omega^*) = 1$ (i.e., $p(q_i^{-1}(t_i),\omega^*) = 1$). Therefore, $\lambda(\cdot,\omega^*) > \mu(\cdot,\omega^*)$ unless $p(\bigcap_{i \in N} q_i^{-1}(t_i),\omega^*) = 1$.

Let $t = (t_1,\ldots,t_n)$. Then,

$$p(\{\omega \in \Omega | q(\omega) = t\}, \omega^*) = 1$$

Combining cases 1 and 2 gives the result that p a.e., $\mathcal{A} \in \Omega$, $\lambda(\cdot,\omega^*) > \mu(\cdot,\omega^*)$ unless q is $p(\cdot,\omega^*)$ a.s. constant.

Let M be the null set (i.e., $p(M) = 0$) on which this result does not hold and recall the definition of regular almost everywhere proper conditional probabilities introduced earlier. There, sets V and V_A were given, with the property that

$$V \in \mathcal{A},\ p(V) = 0 \text{ with } p(G,\omega) = 1 \text{ if } \omega \in G \in \mathcal{A},\ \omega \notin V$$

$$V_A \in \mathcal{A},\ p_A(V_A) = 0 \text{ with } p_A(G,\omega) = 1 \text{ if } \omega \in G \in \mathcal{A},\ \omega \notin V_A$$

observe that since $\mathcal{A} \subseteq \mathcal{F}_i$, $\forall\ i \in N$, $V_A \in \mathcal{F}_i$ so that

$$\int_{V_A} q_i dp = \int_{V_A} p(A|\mathcal{F}_i)dp = p(A \cap V_A)$$

Since $0 = p_A(V_A) = p(A \cap V_A)/p(A)$ and $p(A) > 0$ this implies that

$$\int_{V_A} q_i dp = 0,\ \forall\ i \in N$$

so that on V_A, $q_i = 0$ p a.e.

Let $\omega^* \in \Omega \setminus (M \cup V \cup V_A)$ and let $Q \in \mathcal{A}$ satisfy

$$\omega^* \in Q \subseteq \{\omega \in \Omega | h(q(\omega)) = h(q(\omega^*))\}$$

(Recall that $h(q(\omega)) = \phi(\omega)$ and by assumption $\sigma(\phi) \subseteq \mathcal{A}$). Since $h = g(f)$, where g is invertible on its range, $\forall\ \omega \in Q$, $f(q(\omega)) = f(q(\omega^*))$.

Next,

$$f(\lambda(\cdot,\omega^*)) = \int f(x)\lambda(dx,\omega^*)$$

$$= \int f(q(\omega))p_A(d\omega,\omega^*)$$

and using the fact that $p_A(Q,\omega^*) = 1$, since $\omega^* \notin V_A$ gives
$f(\lambda(\cdot,\omega^*)) = f(q(\omega^*))\int p_A(d\omega,\omega^*) = f(q(\omega^*))$ and since $\omega^* \notin V$,

$$f(q(\omega^*)) = \int f(q(\omega))p(d\omega,\omega^*)$$

$$= \int f(x)\mu(dx,\omega^*) = f(\mu(\cdot,\omega^*))$$

Since f is stochastically monotone, this implies that $\lambda(\cdot,\omega^*) \not\succ \mu(\cdot,\omega^*)$. Thus $\omega^* \in \Omega \setminus (M \cup V \cup V_A)$ and $Q \subseteq \mathcal{A}$ with
$\omega^* \in Q \subseteq \{\omega \in \Omega | h(q(\omega)) = h(q(\omega^*))\}$ implies $\lambda(\cdot,\omega^*) \not\succ \mu(\cdot,\omega^*)$ so that q is $p(\cdot,\omega^*)$ a.s. constant.

Note that $p(V_A)$ may be positive, however

$$0 = \int_{V_A} q_i dp = \int_{V_A} [\int q_i p(d\omega,\omega^*)]p(d\omega^*)$$

so

$$p \text{ a.e. } \omega^* \in V_A, \int q_i p(d\omega,\omega^*) = 0$$

or

$$p \text{ a.e. } \omega^* \in V_A, q_i = 0, p(\cdot,\omega^*) \text{ a.c.}$$

Consequently, p a.e. $\omega^* \in \Omega$, if $\omega^* \in Q \subseteq \{\omega \in \Omega | h(q(\omega)) = h(q(\omega^*))\}$ and $Q \in \mathcal{A}$, then q is constant $p(\cdot,\omega^*)$ a.e. Recalling Lemma 2 with C = I, p a.e. $\omega^* \in \Omega$

$$p(A|\mathcal{A})(\omega^*)\lambda_i(I,\omega^*) = \int_I t\mu_i(dt,\omega^*)$$

or

$$p(A|\mathcal{A})(\omega^*) = \int_I t\mu_i(dt,\omega^*)$$

Using the change of variable formula gives

$$p(A|\mathcal{A})(\omega^*) = \int_{q_i^{-1}(I)} q_i(\omega)p(d\omega,\omega^*)$$

$$= \int q_i(\omega)p(d\omega,\omega^*)$$

For all $\tilde{\omega} \in \Omega$, let $\hat{q}_i(\tilde{\omega}) = \int q_i(\omega)p(d\omega,\tilde{\omega})$. Thus, with p a.e. $\omega^* \in \Omega$

$$p(A|\mathcal{A})(\omega^*) = \hat{q}_i(\omega^*)$$

Since p a.e. $\omega^* \in \Omega$, $q_i(\omega)$ is constant $p(\cdot,\omega^*)$ a.e., it follows that p a.e. $\omega^* \in \Omega$

$$\int |\hat{q}_i(\omega) - q_i(\omega)| p(d\omega,\omega^*) = 0$$

Therefore,

$$\int [\int |\hat{q}_i(\omega) - q_i(\omega)| p(d\omega,\omega^*)] p(d\omega^*) = 0$$

or

$$\int |\hat{q}_i(\omega) - q_i(\omega)| p(d\omega) = 0$$

so that $\hat{q}_i = q_i$, p a.e., and so

$$p(A|\mathcal{A}) = q_i, \text{ p a.e.}$$

This completes the proof of Proposition 1.

Returning to the proof of Theorem 2, observe that

$$q_i^t = p(A|\mathcal{F}_i \vee \mathcal{A}^{t-1}),\ i \in N$$

Thus, $(q_i^t, \mathcal{F}_i \vee \mathcal{A}^{t-1})$ is a martingale so that for all $i \in N$, q_i^t converges almost surely to $p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)$, where $\mathcal{A}^\infty = \bigvee_{s\geq 1} \mathcal{A}_s$.

Let G be a set such that $p(G) = 0$ and $\forall\ i \in N$,

$$q_i^t(\omega) \to p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)(\omega),\ \forall\ \omega \in G^c$$

Note that $\phi^t = h(q^t)$ and since $\forall\ i \in N$, q_i^t converges on G^c and h is continuous, ϕ^t converges on G^c. Let $\underline{\phi} = \liminf \phi^t$, so $\underline{\phi}$ is an $\mathcal{A}^\infty$ measurable function with

$$\underline{\phi}(\omega) = \lim \phi^t(\omega) = h(\{p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)(\omega)\}_{i\in N}),\ \omega \in G^c$$

Pick some $\bar{q} \in I^n$ and define q_i^* on Ω as

$$q_i^* = \bar{q}_i I_G + I_{G^c} p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)$$

Observe that q_i^* is an $\mathcal{F}_i \vee \mathcal{A}^\infty \vee \sigma(G)$ measurable function and that q_i^* is a version of $p(A|\mathcal{F}_i \vee a^\infty \vee \sigma(G))$. Next, define ϕ^∞ as

$$\phi^\infty = I_G h(\bar{q}) + I_{G^c}\, \underline{\phi}$$

$$= h(\bar{q} I_G + I_{G^c}\{p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)\}_{i\in I})$$

$$= h(q^*)$$

Note that ϕ^∞ is $\mathcal{A}^\infty \vee \sigma(G)$ measurable (and is an almost sure limit of ϕ^t).

Let $\mathcal{F}_i^1 = \mathcal{F}_i \vee \mathcal{A}^\infty \vee \sigma(G)$, $i \in N$ and let $\mathcal{A}^1 = \mathcal{A}^\infty \vee \sigma(G)$.

Note that $\sigma(\phi^\infty) \subseteq \mathcal{A}^1 \subseteq \bigcap_{i\in N} \mathcal{F}_i^1$

Therefore Proposition 1 may be applied with $\mathcal{F}_i^1$ replacing $\mathcal{F}_i$, $\mathcal{A}^1$ replacing $\mathcal{A}$, ϕ^∞ replacing ϕ and q_i^* replacing q_i.

This yields the result that for all $i,j \in N$, $q_i^* = q_j^*$ a.e. To complete the proof, observe that $q_i^* = p(A|\mathcal{F}_i \vee \mathcal{A}^\infty)$ a.e. $\forall\ i \in N$, so that for all $i,j \in N$

$$p(A|\mathcal{F}_i \vee \mathcal{A}^\infty) = p(A|\mathcal{F}_j \vee \mathcal{A}^\infty) \quad \text{a.e.}$$

Footnote

[1]This suggests the following way of defining a sigma algebra to represent a partition. Let π be a partition of Ω, and let

$$\mathcal{A} = \{x \mid x: \Omega \to \mathbb{R},\ x \text{ measurable with respect to } \mathcal{F} \text{ and } x \text{ constant on elements of } \pi\}$$

Take $\mathcal{F}_\pi = \bigvee_{x \in \mathcal{A}} \sigma(x)$ be the sigma algebra representing π.

Thus, $\mathcal{A}$ is the set of (Borel) measurable functions which are constant on elements of the partition. An element of $\mathcal{A}$ may be interpreted as the decision function of an agent with the partition π. Constancy of any decision function on the elements of π is a necessary requirement so that $\mathcal{A}$ may be viewed as the class of admissible decision functions. $\mathcal{F}_\pi$ is then the smallest sigma algebra with respect to which the class of admissible decision functions is measurable.

References

Blackwell, D. (1955), "On a Class of Probability Spaces," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press.

Blackwell, D. and Ryll-Nardzewski (1963), "Non-existence of Everywhere Proper Conditional Distributions," Annals of Mathematical Statistics, 34, 223-225.

Blackwell, D. and Dubins, L. (1975), "On Existence and Nonexistence of Proper, Regular, Conditional Distributions," The Annals of Probability, 3, 753-761.

Brandenburger, A. and Dekel, E. (1985), "Hierarchies of Beliefs and Common Knowledge," Research Paper No. 841, Department of Economics, Harvard University.

Geanakoplos, J. D. and H. M. Polemarchakis (1982), "We Can't Disagree Forever," Journal of Economic Theory, 28, 192-200.

McKelvey, R. D. and T. Page (1986), "Common Knowledge, Consensus, and Aggregate Information," Econometrica, 54, 109-127.

Nielsen, L. T. (1984), "Common Knowledge, Communication, and Convergence of Beliefs," Mathematical Social Sciences, 8, 1-14.