Discussion Paper No. 983

WEAK AND STRONG MERGING OF OPINIONS*

by

Ehud Kalai**

and

Ehud Lehrer***

February 1992
Revised January 1993

*This is an extended version of our paper, "Merging of Opinions Revisited," 1991. This research was partly supported by NSF Grants Nos. SES-9011790 and SES-9022305, Economics.

**Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management and Department of Mathematics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208.

***School of Mathematics, Tel Aviv University, Tel Aviv, Israel.

[L6-8]

Abstract

We study merging, in a few senses, of two measures when increasing sequence of information is observed. Motivating this extension of Blackwell and Dubins' (1962) work, are studies of convergence to equilibrium in infinite games and in dynamic economies.

WEAK AND STRONG MERGING OF OPINIONS*

by

Ehud Kalai**

and

Ehud Lehrer**-

February 1992

Revised January 1993

*This is an extended version of our paper, "Merging of Opinions Revisited," 1991. This research was partly supported by NSF Grants Nos. SES-9011790 and SES-9022305, Economics.

**Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management and Department of Mathematics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208.

-School of Mathematics, Tel Aviv University, Tel Aviv, Israel.

[L6-8]

1. Introduction and Conventions

If μ is a true probability distribution over a set of possible states, and $\tilde{\mu}$ is a false distribution over the same space, then posterior probabilities computed according to μ and $\tilde{\mu}$ merge in the norm topology as enough information becomes available. Blackwell and Dubins (1962) proved the above result under the assumption that μ is absolutely continuous with respect to $\tilde{\mu}$. Diaconis and Freedman (1986) investigated the weak star convergence of conditional probabilities; Schervish and Seidenfeld (1990) treated compact sets of mutually absolutely continuous measures.

Recent papers by Kalai and Lehrer (1993a and 1990) applied a version of the Blackwell-Dubins' result to illustrate convergence to Nash equilibrium by subjectively rational players engaged in a repeated game, and convergence to rational expectation equilibrium by subjectively rational agents interacting in a dynamic economy.

Section 2 of this paper reviews and expands a version of the Blackwell-Dubins model and results, and gives an alternative characterization of their notion of merging. This alternative characterization, reported earlier in Kalai and Lehrer (1993a), shows more explicitly the power to predict long horizons of games and economies resulting from Blackwell-Dubins' notion of merging. In addition, this section contains a converse result showing that merging and absolute continuity are equivalent notions.

In the repeated game example, a player wishing to maximize his expected utility, starts with a subjective assessment regarding his opponents' strategies. This assessment induces a subjective probability distribution, $\tilde{\mu}$, on the future play of the game, while the true probability distribution, μ, induced by the strategies actually chosen, is likely to be different.

However, as stated above, under an absolute continuity condition and after sufficiently long play, $\tilde{\mu}$ and μ merge. This merging was shown to imply that after a long time the actual play is ε-close to a play induced by some ε-Nash equilibrium of the repeated game.

The above mentioned game theoretic result motivates important mathematical questions. First, the absolute continuity assumption of μ with respect to $\tilde{\mu}$ is strong for some game theoretic and economic applications. At the same time the notion of merging obtained is stronger than needed for such applications. More specifically, this notion of merging implies closeness of μ and $\tilde{\mu}$ even for events in the infinite future. For instance, the event that in the infinite future the long run rate of cooperation is exactly 50 percent.

In economic applications involving discounting of future payoffs, it suffices to approximate events in finite horizons. Thus, one hopes to be able to weaken the absolute continuity assumption even if it yields a weaker, yet sufficient, notion of merging. Such a condition is offered in Section 3 where the appropriate notion of weak merging is studied.

A question of an opposite nature is also motivated by the earlier conclusion that players learn to play ε-close to an ε-Nash equilibrium. Since the Nash equilibrium concept in the above statement is already perturbed, can we obtain a result where the actual play coincides exactly with an ε-Nash equilibrium? For this to occur, it turns out that a stronger notion of merging is needed. We refer to it as strong merging and study it in Section 4.

A third issue motivated by economic applications is the relation of merging to the exact specification of how and what information is obtained.

While, as shown in Section 2, usual merging does not depend on this specification, weak merging is sensitive to it. Given two information sequences for the same σ-field, it is possible that weak merging is obtained in one but not in the other.

Dependencies, of merging, on information sequences are important for the study of more sophisticated economic applications. The game theoretic example discussed before assumes perfect monitoring of players' actions. This means a very detailed information sequence which is often not available due to differential imperfect information that agents obtain. Thus, when studying weak merging in more general repeated games models, for example, the precise information structure has to be considered in order to decide whether merging will occur. All three sections (2, 3, and 4) study the relation of the corresponding merging concepts to specific information sequences.

For the remainder of this paper, we let $(\Omega,\mathcal{B})$ be a measurable space of outcomes, interpreted as the set of states of nature. It is assumed throughout that $\mathcal{B}$ is generated by countably many sets. We fix two probability distributions on $(\Omega,\mathcal{B})$, μ and $\tilde{\mu}$, describing respectively the true distribution over states, and the (possibly false) subjective beliefs of some agent. All unspecified probability statements in the paper are made relative to the true distribution, μ.

A sequence of finite or countable partitions of Ω will describe a dynamic information structure that the agent may possess. So, if the information sequence is $\{\mathcal{P}_t\}$ and the state of nature is $\omega \in \Omega$, then at time t the agent is told $P_t(\omega)$, the element of the partition $\mathcal{P}_t$ that contain ω. Given this information, the Bayesian agent adopts $\tilde{\mu}(\cdot|P_t(\omega))$ as his new distribution over states, while the true distribution now is $\mu(\bullet|P_t(\omega))$. Closeness of these two random measures will be our major interest.

2. Blackwell and Dubins Expanded

We restrict ourselves, throughout the paper, to partition sequences of the following type.

Definition 1: An information sequence (for $(\Omega,\mathcal{B})$ μ and $\tilde{\mu}$) is a sequence of partitions $\{\mathcal{P}_t\}$ satisfying:

(1) $\mathcal{P}_{t+1}$ refines $\mathcal{P}_t$.

(2) With $\mathcal{F}_t$ denoting the field generated by $\mathcal{P}_t$ and $\mathcal{F} = \vee\, \mathcal{F}_t$ denoting the σ-field generated by all $\mathcal{F}_t$, we require that $\mathcal{F} = \mathcal{B}$.

(3) If an element of the partition $\mathcal{P}_t$ is assigned a positive probability by μ, then it is assigned a positive probability by $\tilde{\mu}$.

Condition (1) asserts that information is cumulative. Condition (2) requires that the events that can be discussed are exactly the ones that can be formulated in the language of the information sequence. And Condition (3) states that there is zero probability that the agent ever be surprised by being told something he considered impossible. Technically, this condition guarantees that the agent can perform Bayesian updating with probability one.

An interesting economic example is the evolution of price paths in a dynamic economy. Here an outcome $\omega \in \Omega$ describes an infinite sequence of price vectors with $P_t(\omega)$ describing the history of price vectors up to time t (see Kalai-Lehrer, 1990). $\tilde{\mu}(\bullet|P_t(\omega))$ being close to $\mu(\bullet|P_t(\omega))$ means that the agent's forecast of future prices, given past prices up to time t, is almost accurate.

In a game theoretic example (see Kalai-Lehrer, 1993a), $\omega = (a^1, a^2, \ldots)$ describes an infinite sequence of action-vectors in an infinitely repeated game. In this example, $P_t(\omega)$ describes the history of past action vectors up to time t, $\mu(\bullet | P_t(\omega))$ describe the probability distribution over future actions induced by the actual strategies of the players, and $\tilde{\mu}(\bullet | P_t(\omega))$ describe what an observer believes the distribution of future actions to be.

But different information sequences may be of interest. In a two player version of the example above, an agent may be fully informed at time t of all the past actions of player one, but may be informed, with a delay of d periods, of the actions of player two. An information sequence $\bar{\mathcal{P}} = \{\bar{\mathcal{P}}_t\}$ describing such a process has

$$\bar{\mathcal{P}}_t(a^1, a^2, \ldots) = \{(b^1, b^2, \ldots);\ b_1^j = a_1^j \ j = 1, \ldots, t,\ b_2^j = a_2^j \ j = 1, \ldots, t - d\}.$$

It is easy to see that $\mathcal{P}$ and $\bar{\mathcal{P}}$ are both information sequences for the same σ-algebra of infinite action paths.

The next theorem is a version of the seminal Blacwell and Dubins (1962) result about merging of measures. In order to make the reading of the current paper self contained, we present here a variation of their definitions and proof presented in Kalai and Lehrer (1993a).

<u>Definition 2</u>: For $\varepsilon > 0$ we say that $\tilde{\mu}$ is <u>ε-close to</u> μ if there exists a set Q satisfying

(i) $\mu(Q), \tilde{\mu}(Q) \geq 1 - \varepsilon$, and

(ii) for every event $A \subset Q$.

(1) $(1 - \varepsilon)\tilde{\mu}(A) \le \mu(A) \le (1 + \varepsilon)\tilde{\mu}(A)$.

We say that $\tilde{\mu}$ merges to μ in the information sequence $\{\mathcal{P}_t\}$ if for every $\varepsilon > 0$ and almost every ω there is a time $r(\varepsilon,\omega)$ such that for every $r \ge r(\varepsilon,\omega)$ $\tilde{\mu}(\cdot|P_r(\omega))$ is ε-close to $\mu(\cdot|P_r(\omega))$. We say that $\tilde{\mu}$ merges to μ if it does so for every information sequence. (Merging, and merging in some information sequence, will turn out to be equivalent; see Corollary 1.)

The notion of closeness described in (1) is strong. It implies that even for small probability events in Q, μ can differ from $\tilde{\mu}$ only by a small percentage. Also, for A and B in Q, conditional probabilities of A given B computed according to μ can differ from those computed according to $\tilde{\mu}$ only by a small percentage. This shows that once closeness in the sense of (1) was obtained, the agent's probabilities and conditional probabilities (for events in Q) will always be approximately correct without building cumulative mistakes.

Proposition 1: Let $\{\mathcal{P}_t\}$ be an information sequence. Suppose that μ is absolutely continuous with respect to $\tilde{\mu}$, $\mu << \tilde{\mu}$ (formally, $\mu(A) > 0$ implies $\tilde{\mu}(A) > 0$ for every $A \in \mathcal{B}$). With μ-probability 1 for every $\varepsilon > 0$ there is a random variable $t(\varepsilon,\omega)$ s.t. for every $s \ge t \ge t(\varepsilon,\omega)$.

(2) $$1 - \varepsilon \le \frac{\mu(P_s(\omega)|P_t(\omega))}{\tilde{\mu}(P_s(\omega)|P_t(\omega))} \le 1 + \varepsilon.$$

Proof: Since $\mu << \tilde{\mu}$, by the Radon-Nikodym theorem, there is an $\mathcal{F}$-measurable function f satisfying

(3) $\int_A f d\tilde{\mu} = \mu(A)$ for every $A \in \mathcal{F}$.

By Levy's theorem (see Shiryayev, 1984), $E_{\tilde{\mu}}(f|\mathcal{F}_t) \to E_{\tilde{\mu}}(f|\mathcal{F}) = f$ $\tilde{\mu}$ almost surely (and therefore, μ-a.s.). However, for $\tilde{\mu}$ almost all ω

(4) $$E_{\tilde{\mu}}(f|\mathcal{F}_t)(\omega) = 1/\tilde{\mu}(P_t(\omega)) \int_{P_t(\omega)} f d\tilde{\mu} = \mu(P_t(\omega))/\tilde{\mu}(P_t(\omega)).$$

Moreover, by (3), $f > 0$ μ-a.s. Thus, the right side of (4) tends μ-a.s. to a positive number. In other words, there is a $t(\varepsilon)$ such that for μ almost all ω the following holds:

(5) $$1 - \varepsilon \le \frac{\mu(P_t(\omega))}{\tilde{\mu}(P_t(\omega))} \Big/ \frac{\mu(P_s(\omega))}{\tilde{\mu}(P_s(\omega))} \le 1 + \varepsilon \text{ for all } s \ge t \ge t(\varepsilon).$$

The middle term of (5) is equal to the middle one in (2). Since (2) holds for every $\varepsilon > 0$ with probability 1, the proposition follows. //

Remark 1: Proposition 1 can be proved directly by using the martingale convergence theorem and without the Radon-Nikodym theorem (which is a by-product) as follows. Define $X_t(\omega) = \mu(P_t(\omega))/\tilde{\mu}(P_t(\omega))$. Obviously, this is $\tilde{\mu}$-martingale and thus converges $\tilde{\mu}$-a.s. to X_∞. Furthermore, $\mu << \tilde{\mu}$ on $\mathcal{F}$ if and only if $\{X_t\}$ is uniformly integrable--i.e., $\lim_{c\to\infty} \sup_t \int_{\{X_t>c\}} X_t d\tilde{\mu} = 0$. Thus, $\int_A X_\infty d\tilde{\mu} = \int_A X_t d\tilde{\mu} = \mu(A)$ for every $A \in \mathcal{F}_t$. Therefore, $\int_A X_\infty d\tilde{\mu} = \mu(A)$ for every $A \in \mathcal{F}$. Since $\mu << \tilde{\mu}$, X_∞ is positive μ - a.s.

Theorem 1 (see Blackwell and Dubins, 1962, for an earlier version): If

$\mu << \tilde{\mu}$ then $\tilde{\mu}$ merges to μ.

The proof follows from Proposition 1 and the following lemma:

Lemma 1: Let g_t be a sequence of measurable functions which converges μ a.s. to $k \neq 0$. For every $\varepsilon > 0$ and $d > 0$ there is a time, $t(\varepsilon,d)$, s.t. $\mu\{\omega;\mu(C_t|P_t(\omega)) > \varepsilon$ for at least one $t \geq t(\varepsilon,d)\} < \varepsilon$, where $C_t = \{\omega; |g_s(\omega)/k(\omega) - 1| > d, s \geq t\}$.

Proof: Since $g_t \to k \neq 0$, the sequence $\mu(C_t)$ converges to 0. Suppose to the contrary that the lemma does not hold. Then there is a μ-positive set A, $d > 0$ and $\varepsilon > 0$ such that for all $\omega \in A$, $\mu(C_t|P_t(\omega)) > \varepsilon$ for infinitely many t's.

Fix $s \in \mathbb{N}$ and define

$$B_r = \{\omega \in A; r = \min\{t; t \geq s \text{ and } \mu(C_t|P_t(\omega)) > \varepsilon\}\}.$$

Observe that $\{B_r\}$ are pairwise disjoint and, moreover, $\{\cup_{\omega\in B_r} P_r(\omega)\}_r$ are also pairwise disjoint. By the definition, $A = \cup_{r\geq s} B_r$.

Since $C_s \supseteq C_t$ when $t \geq s$, for all $\omega \in A$, $\mu(C_s|P_t(\omega)) > \varepsilon$ for every $\omega \in B_t$. Thus, $\mu(C_s|\cup_{\omega\in B_t} P_t(\omega)) > \varepsilon$. Therefore, $\mu(C_s) > \varepsilon\cdot\mu(\cup_{t\geq s} \cup_{\omega\in B_t} P_t(\omega)) \geq \varepsilon\cdot\mu(\cup_{t\geq s} B_t) = \varepsilon\cdot\mu(A)$.

Hence, the sequence $\{\mu(C_s)\}$ is bounded away from zero, which is a contradiction. //

In order to complete the proof of Theorem 1, apply Lemma 1 to g_t = the indicator function of the set $\{\omega; |E(f|\mathcal{F}_s)(\omega)/E(f|\mathcal{F}_t)(\omega) - 1| < \varepsilon$ for all $s \geq t\}$. Obviously, $g_t \to 1$.

Remark 2: Blackwell and Dubins use a notion of closeness of measures stated as a bound on the absolute differences of probabilities, i.e.,

$$(6) \qquad |\mu(A|P_t) - \tilde{\mu}(A|P_t)| < \varepsilon \text{ for every event } A \subseteq P_t.$$

Our notion of closeness, inequality (1) in Definition 2, can be rewritten as $|\mu(A|P_t) - \tilde{\mu}(A|P_t)| \leq \varepsilon\tilde{\mu}(A|P_t)$ for all events A in a large set. So it is easy to see that ε-closeness according to our definition implies 3ε-closeness in the snese of Blackwell and Dubins. The following proposition, however, illustrates that a converse is also true and thus the two notions are asymptotically equivalent. Thus Theorem 1 is equivalent to a version of the original Blackwell and Dubins' statement.

Proposition 2:

(i) for every $\varepsilon > 0$ there exists $\delta > 0$ s.t. if $|\mu(A) - \tilde{\mu}(A)| < \delta$ for every event A, then $\tilde{\mu}$ is ε-close to μ;

(ii) if $\tilde{\mu}$ is ε-close to μ then $|\mu(A) - \tilde{\mu}(A)| < 3\varepsilon$ for every event A.

Proof:

(i) Let $1/4 > \delta > 0$, to be specified later, and assume that $\mu, \tilde{\mu}$ satisfy $|\mu(A) - \tilde{\mu}(A)| < \delta$ for every event A.

By using the Lebesgue decomposition theorem (see Halmos, 1950, p. 143), we will define three events, S, S_1, and S_2 which satisfy:

(a) for every $\Phi \neq D \subseteq S_1$, $\mu(D) > 0$, $\tilde{\mu}(D) = 0$;

(b) for every $\Phi \neq D \subseteq S_2$, $\tilde{\mu}(D) > 0$, $\mu(D) = 0$;

(c) $S \cup S_1 \cup S_2 = \Omega$.

Thus, on S, μ and $\tilde{\mu}$ are mutually absolutely continuous.

The measures μ and $\tilde{\mu}$ can be decomposed (by the Lebesgue decomposition theorem) as follows: $\mu = \mu^1 + \mu^2$ and $\tilde{\mu} = \tilde{\mu}^1 + \tilde{\mu}^2$, where[1] $\mu^1 << \tilde{\mu}$, $\mu^2 \perp \tilde{\mu}$ and $\tilde{\mu}^1 << \mu$, $\tilde{\mu}^2 \perp \mu$. By the definition of singularity there are two sets S_1 and S_2 satisfying $\mu^1(D \cap S_1) = \mu^1(D)$, $\tilde{\mu}(D \cap S_1^c) = \tilde{\mu}(D)$, $\tilde{\mu}^1(D \cap S_2) = \tilde{\mu}^1(D)$, and $\mu(D \cap S_2^c) = \mu(D)$ for every D. Hence, S_1 and S_2 satisfy (a) and (b). Define $S = \Omega \backslash (S_1 \cup S_2)$. $\mu(S_1 \cup S_2) = \mu(S_1)$ and $\tilde{\mu}(S_1) = 0$. Thus, $\mu(S_1 \cup S_2) < \delta$. For a similar reason, $\tilde{\mu}(S_1 \cup S_2) < \delta$. Therefore, $\mu((S_1 \cup S_2)^c)$ and $\tilde{\mu}((S_1 \cup S_2)^c)$ are at least $1 - \delta$.

By the Radon-Nikodym theorem there exists a measurable function f satisfying

$$\mu(D) = \int_D f d\tilde{\mu} \text{ for all events } D \subseteq S.$$

Define $\bar{B} = \{\omega | f(\omega) - 1 > \sqrt{\delta}\}$, and

$\underline{B} = \{\omega | f(\omega) - 1 < -\sqrt{\delta}\}$.

Observe that $\mu(\bar{B}) - \tilde{\mu}(\bar{B}) > \sqrt{\delta}\tilde{\mu}(\bar{B})$. Therefore, $\tilde{\mu}(\bar{B}) < \sqrt{\delta}$. For a similar reason, $\tilde{\mu}(\underline{B}) < \sqrt{\delta}$. Defining $Q = S - (\bar{B} \cup \underline{B})$, we get $\tilde{\mu}(Q) \geq 1 - \delta - 2\sqrt{\delta}$. Moreover, for every event $D \subseteq Q$ one gets $|\mu(D)/\tilde{\mu}(D) - 1| < 2\sqrt{\delta}$. Therefore, if $\delta + 2\sqrt{\delta} < \varepsilon$, then μ is ε-close to $\tilde{\mu}$.

---

[1] Two measures, λ_1, λ_2 are <u>singular</u> ($\lambda_1 \perp \lambda_2$) if there is a set A s.t. for every event D, $\lambda_1(D \cap A) = \lambda_1(D)$ and $\lambda_2(D \cap A^c) = \lambda_2(D)$.

(ii) If μ is ε-close to $\tilde{\mu}$, then there is a set Q s.t. $\mu(Q)$, and $\tilde{\mu}(Q)$ are greater than $1 - \varepsilon$, and moreover for every $A \subseteq Q$ $(1 - \varepsilon)\tilde{\mu}(A) \le \mu(A) \le (1 + \varepsilon)\mu(A)$. Thus, for an event B, $\tilde{\mu}(B) = \tilde{\mu}(B \cap Q) + \tilde{\mu}(B \cap Q^c) \le (1/(1 - \varepsilon))\mu(B \cap Q) + \varepsilon \le (1/(1 - \varepsilon))\mu(B) + \varepsilon$ and $\mu(B) = \mu(B \cap Q) + \mu(B \cap Q^c) \le (1 + \varepsilon)\tilde{\mu}(B \cap Q) + \varepsilon \le (1 + \varepsilon)\tilde{\mu}(B) + \varepsilon$. Combining these two inequalities one gets the desired proof of (ii). //

Next we prove a strong converse to Theorem 1.

<u>Theorem 2</u>: If $\tilde{\mu}$ merges to μ in some information sequence then $\mu << \tilde{\mu}$.

<u>Proof</u>: Assuming that $\tilde{\mu}$ merges to μ in the information sequence $\{\mathcal{P}_t\}$, suppose, contrary to the statement of the theorem, that for some event A, $\mu(A) > 0$ and $\tilde{\mu}(A) = 0$. Since $A \in V\mathcal{P}_t$ for every $\varepsilon > 0$ there is t arbitrarily large and $B_1^t, \dots, B_{k_t}^t \in \mathcal{P}_t$ s.t. $\mu(\cup_{i=1}^{k_t} B_i^t \,\Delta\, A) < \varepsilon\mu(A)$. Thus, $\mu(A|\cup_{i=1}^{k_t} B_i^t) \ge (1 - \varepsilon)\mu(A)/(1 + \varepsilon)\mu(A) = (1 - \varepsilon)/(1 + \varepsilon) \ge 1 - 2\varepsilon$ and $\mu(\cup_{i=1}^{k_t} B_i^t) > (1 - \varepsilon)\mu(A)$. Note that $\mu(A|\cup_{i=1}^{k_t} B_i^t) = \Sigma_{i=1}^{k_t} (A|B_i^t)[\mu(B_i^t)/\mu(\cup B_i^t)]$. In particular, $\mu(A|\cup_{i=1}^{k_t} B_i^t)$ is a convex combination of $\mu(A|B_i^t)$, $t = 1, \dots, k_t$. Thus, the μ-measure of all those B_i^t s.t. $\mu(A|B_i^t) \ge 1 - \sqrt{\varepsilon} > 1/2$ is at least $(1 - 2\sqrt{\varepsilon})\mu(\cup B_i^t)$. Therefore, the μ-measure of these B_i^t's is at least $(1 - 2\sqrt{\varepsilon})(1 - \varepsilon)\mu(A) \ge (1 - 3\sqrt{\varepsilon})\mu(A)$.

By taking a fast converging sequence of ε's (e.g., $1/n^4$) one can find an event C satisfying (i) $\mu(C) \ge (1/2)\mu(A)$ and (ii) for every $\omega \in C$ there are infinitely many t's s.t. $\mu(A|P_t(\omega)) > 1/2$. By our assumption,

$\tilde{\mu}(A|P_t(\omega)) = 0$. These contradict (ii) of Definition 2. //

Combining Theorem 2 with Theorem 1 we obtain

<u>Corollary 1</u>: For any two information sequences $\mathcal{P}$ and $\bar{\mathcal{P}}$, $\tilde{\mu}$ merges to μ in $\mathcal{P}$ iff $\tilde{\mu}$ merges to μ in $\bar{\mathcal{P}}$.

<u>Remark 3</u>: Corollary 1 shows that we can equivalently define merging by requiring only that it occurs in some information sequence. Thus, $\tilde{\mu}$ merges to μ vacuously, if there are no information sequences, or if it merges in at least one information sequence.

The above observation shows, together with Lemma 2 below, that we have the following alternative characterization of absolute continuity: the probability measure μ is absolutely continuous w.r.t. $\tilde{\mu}$ iff $\tilde{\mu}$ merges to μ in some information sequence.

Necessary and sufficient conditions for the existence of an information sequence are given in the following lemma. The measure $\tilde{\mu}$ is said to be <u>finitely atomic</u> if there are finitely many atoms, $A_1,\ldots,A_n$, of $\tilde{\mu}$, s.t., $\Sigma_1^n \tilde{\mu}(A_i) = 1$.

<u>Lemma 2</u>: There is an information sequence iff either (i) $\mu \ll \tilde{\mu}$, or (ii) $\tilde{\mu}$ is not finitely atomic.

<u>Proof</u>: Suppose first that there exists an information sequence. If (i)

does not hold, then there is an event A s.t. $\mu(A) > 0$ and $\tilde{\mu}(A) = 0$. If $\{\mathcal{P}_t\}$ is an information sequence and $\{\mathcal{F}_t\}$ is the corresponding sequence of fields then there is a sequence $B_t \in \mathcal{F}_t$ $(\mu + \tilde{\mu})(A \Delta B_t) \to 0$. Thus, $\mu(B_t) > 0$ and, therefore, $\tilde{\mu}(B_t) > 0$. Moreover, $\tilde{\mu}(B_t) \to 0$. Therefore, $\tilde{\mu}$ is not finitely atomic.

To prove the converse direction assume first that (i) is satisfied. It is clear that since $\mathcal{B}$ is countably generated, there is an information sequence.

Suppose now that (i) is not satisfied but (ii) is. Thus, we may assume that there exists a sequence of pairwise disjoint sequences of events $\{D_n\}$ s.t. $\tilde{\mu}(D_n) \to 0$.

Since $\mathcal{B}$ is countably generated there is a sequence $A_1, A_2, \ldots$ of events which generates $\mathcal{B}$. We will construct inductively an information sequence $\{Q_t\}$.

Let $\{E_i\}$ be a sequence of events in which each one of the A_i's appears infinitely many times.

During the construction we will ensure that each one of the atoms of $\{Q_t\}$ will contain infinitely many D_i's. Therefore, the $\tilde{\mu}$-probability of each one of these atoms is positive. This will take care of (3) of Definition 1.

Let Q_0 be the trivial partition. Suppose that $\{Q_i\}$ for $i \leq t$ have already been constructed with the following properties:

(P1) Q_{i+1} refines Q_i, $i = 1, \ldots, t - 1$.

(P2) For every $1 \leq i \leq t$, each atom of Q_i contains infinitely many D_j's.

(P3) There is a set, E_i', which is a union of atoms of Q_i which satisfies $\mu(E_i' \Delta E_i) \le 1/i$, $1 < i \le t$.

The partition Q_{t+1} is constructed in two steps.

<u>Step 1</u>: Define R_{t+1} to be the partition generated by the partition Q_t and the event E_{t+1}.

In Step 1, each atom of Q_t is divided into at most two atoms of R_{t+1}. Q_{t+1} is defined by changing the atoms of R_{t+1}, by altering the location of some D_j's or parts of them, so that (P2) will hold for Q_{t+1}.

<u>Step 2</u>: The construction is described for arbitrary atoms of R_{t+1}.

Suppose that Y and Z are atoms of R_{t+1}, where $Y \cup Z$ is an atom of Q_t. Since in $Y \cup Z$ there are infinitely many D_j's, one can (i) shift some of D_j's or parts of them from Y to Z or from Z to Y and (ii) rename these atoms as Y' and Z' so that both Y' and Z' will contain infinitely many D_j's. Moreover, these changes can be arbitrarily small in terms of $\tilde{\mu}$-probability (because $\tilde{\mu}(D_n) \to 0$). After doing the required changes for all the atoms of R_{t+1} we get the partition Q_{t+1}.

By this construction (P1) and (P2) are satisfied also for $i = t + 1$. (P3) is satisfied for $i = t + 1$ if the total probability of all the changes in Step 2 of stage $t + 1$ do not exceed $1/(t + 1)$. Since each one of the A's appears infinitely many times in the sequence $\{E_i\}$ and due to (P3) it follows that $A_j \in VQ_t$ for every j. Thus, $\mathcal{F} = VQ_t$. Moreover, by the construction, each one of the atoms of each Q_t contains a lot of D_j's and therefore its $\tilde{\mu}$-probability is positive. Hence, $\{Q_t\}$ is an information sequence. //

3. Weak Merging of Opinions

In this section we introduce and study weak merging of opinions. For every partition $\mathcal{P}_t$ of Ω we denote by $\sigma(\mathcal{P}_t)$ the field generated by $\mathcal{P}_t$.

Definition 3: Let $\mathcal{P} = \{\mathcal{P}_t\}$ be an information sequence. We say that $\tilde{\mu}$ $\mathcal{P}$-weakly merges to ($\mathcal{P}$-w.m.t.) μ if with μ-probability 1, for every $\varepsilon > 0$ and $\ell \in \mathbb{N}$ there exist $r(\varepsilon,\ell,\omega)$ s.t. for every $r > r(\varepsilon,\ell,\omega)$

$$(7) \qquad |\mu(A|P_r(\omega)) - \tilde{\mu}(A|P_r(\omega)| < \varepsilon \text{ for every } A \in \sigma(\mathcal{P}_{r+\ell}).$$

Notice that (7), as opposed to (6), requires the inequality to hold only if A belongs to a finite horizon future.

Remark 4: Since $\mu(A|C)\mu(C|B) = \mu(A|B)$ whenever $C \subseteq B$, the previous definition can be rewritten with $\ell = 1$. Also, given our interest in discounting future values, the following equivalent characterization of $\mathcal{P}$-weak merging may be of interest: for every $0 < \delta < 1$ with μ-probability 1 for every $\varepsilon > 0$ there exists $r(\varepsilon,\delta,w)$ s.t. for every $r > r(\varepsilon,\delta,w)$, $\Sigma_{t\geq r}\, \delta^{t-r} |\mu(A_t|P_r(\omega)) - \tilde{\mu}(A_t|P_r(\omega))| < \varepsilon$ for every sequence of events $A_r, A_{r+1}, \ldots$ with $A_{r+j} \in \sigma(\mathcal{P}_{r+j})$.

The following example shows that, unlike merging, weak merging is indeed partition dependent.

Example: Suppose that the measure space Ω is $\{0,1\}^{\infty}$ and that μ is the probability measure induced on it by the repeated toss of a fair coin. Let

$\tilde{\mu}$ be the probability measure generated by the repeated toss of a biased coin with parameter λ chosen according to the uniform distribution on $[0,1]$.

If $\mathcal{P}_t$ is the partition corresponding to the first t outcomes, then $\tilde{\mu}$ $\{\mathcal{P}_t\}$-weakly merges to μ. However, if $\mathcal{P}_t'$ is the partition corresponding to the 2^{2^t} first observations, then $\tilde{\mu}$ does not $\{\mathcal{P}_t'\}$-weakly merge to μ. The difficulty with $\{\mathcal{P}_t'\}$ is that looking one step ahead in $\mathcal{F}_t'$ requires looking L steps into the "real" future with an exponentially increasing L.

We start with the following simple observation. Recall that $X_t(\omega) = \mu(P_t(\omega))/\tilde{\mu}(P_t(\omega))$. Define $Y_t = |X_{t+1}/X_t - 1|$ and denote $Y_t^c = \min\{Y_t, c\}$ for every positive number c.

Proposition 3:

a. If $X_{t+1}/X_t \to 1$ μ-a.s., then $\tilde{\mu}$ $\{\mathcal{P}_t\}$-weakly merges to μ, and the converse is true if the series of the conditional probabilities, $\{\tilde{\mu}(P_{t+1}(\omega)|P_t(\omega))\}_t$, is bounded away from zero μ-a.s.

b. If $\tilde{\mu}$ $\{\mathcal{P}_t\}$-w.m.t. μ then X_{t+1}/X_t converges in probability (w.r.t. μ) to 1.

c. $\tilde{\mu}$ $\{\mathcal{P}_t\}$-weakly merges to μ iff for every $c > 0$, $E_\mu(Y_t^c|\mathcal{P}_t) \to 0$ μ-a.s. as t goes to infinity.

Proof:

a. Since $X_{t+1}/X_t \to 1$ μ-a.s. it follows that

$$(\mu(P_{t+1}(\omega))/\tilde{\mu}(P_{t+1}(\omega))/(\mu(P_t(\omega))/\tilde{\mu}(P_t(\omega)) =$$
$$= \mu(P_{t+1}(\omega)|P_t(\omega))/\tilde{\mu}(P_{t+1}(\omega)|P_t(\omega)) \to 1 \ \mu \text{ a.s.}$$

We use now Lemma 1 to derive that there is a random variable $t(\varepsilon,\omega)$ s.t. if $t > t(\varepsilon,\omega)$ then for μ almost all ω

$$\mu\{\omega';\ P_t(\omega') = P_t(\omega) \text{ and } \left|\frac{\mu(P_{t+1}(\omega')|P_t(\omega))}{\tilde{\mu}(P_{t+1}(\omega')|P_t(\omega))} - 1\right| > \varepsilon|P_t(\omega)\} < \varepsilon,$$

which shows that $\tilde{\mu}$ $\{\mathcal{P}_t\}$-w.m.t. μ. As for the converse claim, if $\tilde{\mu}(P_{t+1}(\omega)|P_t(\omega))$ are all bounded away from zero μ a.s. then there exists a constant c s.t. whenever

$$|\mu(P_{t+1}(\omega)|P_t(\omega)) - \tilde{\mu}(P_{t+1}(\omega)|P_t(\omega))| < \varepsilon,$$

the following holds:

$$|\mu(P_{t+1}(\omega)|P_t(\omega))/\tilde{\mu}(P_{t+1}(\omega)|P_t(\omega)) - 1| < c\varepsilon.$$

Thus, $|X_{t+1}(\omega)/X_t(\omega) - 1| < c\varepsilon$. Therefore, if $\tilde{\mu}$ $\{\mathcal{P}_t\}$-w.m.t. μ then $X_{t+1}/X_t \to 1$ μ-a.s.

b. Fix an $\varepsilon > 0$. If $|\mu(A|P_t(\omega)) - \tilde{\mu}(A|P_t(\omega))| < \varepsilon$

for every A in $\sigma(\mathcal{P}_{t+1})$ then

$$\mu(X_{t+1}/X_t - 1 > \sqrt{\varepsilon}|P_t(\omega)) < 2\sqrt{\varepsilon}$$

and

$$\mu(1 - X_{t+1}/X_t > \sqrt{\varepsilon} | P_t(\omega)) < 2\sqrt{\varepsilon}.$$

Therefore,

$$\mu(|X_{t+1}/X_t - 1| > \sqrt{\varepsilon} | P_t(\omega)) < 4\sqrt{\varepsilon}.$$

Define $A_T = \{\omega;$ if $t \geq T$ then $|\mu(A|P_t(\omega)) - \tilde{\mu}(A|P_t(\omega))| < \varepsilon$ for every A in $\sigma(\mathcal{P}_{t+1})\}$.

As $\tilde{\mu}$ $\{\mathcal{P}_t\}$-w.m.t. μ, $\mu(A_T) \to 1$ as $T \to \infty$. Therefore, there is a T s.t. $t \geq T$ implies $\mu(A_t) > 1 - \varepsilon$. Thus, if $t \geq T$ then $\mu(|X_{t+1}/X_t - 1| < \sqrt{\varepsilon}) > (1 - \varepsilon)(1 - 4\sqrt{\varepsilon})$. Since ε is arbitrary, it shows that X_{t+1}/X_t converges in probability (w.r.t. μ) to 1.

c. Weak merging of $\tilde{\mu}$ to μ w.r.t. $\{\mathcal{P}_t\}$ is equivalent to saying that for every $\varepsilon > 0$

$$\mu(|X_{t+1}/X_t - 1| > \varepsilon | P_t) \to 0 \quad \mu\text{-a.s.}$$

Since Y_t^c is bounded it is equivalent to $E(Y_t^c|\mathcal{P}_t) \to 0$ μ-a.s. //

<u>Theorem 3</u>: $\tilde{\mu}$ $\{\mathcal{P}_t\}$-w.m.t. μ for every information sequence $\{\mathcal{P}_t\}$ iff $\tilde{\mu}$ merges to μ.

<u>Proof</u>: If there is no information sequence the equivalence is obvious. So we assume that there is an information sequence. According to Theorem 1, if $\tilde{\mu}$ does not merge to μ then μ is not absolutely continuous w.r.t. μ. Thus, there exists an $A \in \mathcal{B}$ s.t. $\mu(A) > 0$ and $\tilde{\mu}(A) = 0$. Now let $\{\mathcal{P}_t\}$ be an information sequence. Define $X_t(\omega)$ as in Remark 1. $\{X_t\}$ is $\tilde{\mu}$ martingale and therefore converges $\tilde{\mu}$-a.s. to X_∞. Since $\{\mathcal{P}_t\}$ satisfies condition (3)

of Definition 1 (see Shiryayev, p. 493, Th. 1):

$$\mu(B) = \int_B X_\infty d\tilde{\mu} + \mu(B \cap \{X_\infty = \infty\})$$

for every $B \in \mathcal{F}$.

Therefore, $X_t \to \infty$ on A. One can form a subsequence of $\mathcal{P}_{t_m}$ s.t. X_{t_m} will converge fast to ∞ in the sense that $X_{t_{m+1}}/X_{t_m} \to \infty$ on A. According to Proposition 3b, $\tilde{\mu}$ does not $\{\mathcal{P}_{t_m}\}$-w.m.t. μ. //

4. Strong Merging of Opinions

In this section we introduce and study a notion of strong merging.

Definition 4: Let $\{\mathcal{P}_t\}$ be an information sequence. $\tilde{\mu}$ $\{\mathcal{P}_t\}$-strongly merges to (s.m.t.) $\tilde{\mu}$ if with probability one for every $\varepsilon > 0$ there exists $t(\varepsilon,\omega)$ s.t. if $t > t(\varepsilon,\omega)$ then

$$|\mu(A|P_t(\omega))/\tilde{\mu}(A|P_t(\omega)) - 1| < \varepsilon \text{ for every measurable } A \subseteq P_t(\omega).$$

It is clear that if $\tilde{\mu}$ $\{\mathcal{P}_t\}$-s.m.t. μ then $\mu << \tilde{\mu}$. Let $f = d\mu/d\tilde{\mu}$, the Radon-Nikodym derivative, and denote[2] for any $\{\mathcal{P}_t\}$

$$Y_t(\omega) = \sup\{f(x)/f(y); x,y \in P_t(\omega)\}.$$

Theorem 4: $\tilde{\mu}$ $\{P_t\}$-s.m.t. μ iff $Y_t(\omega) \to 1$ μ-a.s.

---

[2]The supremum and the infimum in the sequel are modulo sets of measure 0.

<u>Proof</u>: Suppose first that $Y_t(\omega) \to 1$. Now, for every $A \subseteq P_t(\omega)$

$$\inf\{f(x);\ x \in P_t(\omega)\}\tilde{\mu}(A) \leq \mu(A) \leq \sup\{f(x);x \in P_t(\omega)\}\tilde{\mu}(A).$$

Therefore,

$$(8) \qquad 1/Y_t(\omega) \leq \mu(A|P_t(\omega))/\tilde{\mu}(A|P_t(\omega)) \leq Y_t(\omega).$$

However, $Y_t(\omega)$ converges μ-a.s. to 1. Hence, the middle term of (8) converges to 1 for every $A \subseteq P_t(\omega)$.

If $Y_t(\omega)$ does not converge μ-a.s. to 1, then there exist a set A and $\varepsilon > 0$ satisfying: (i) $\mu(A) > 0$ and (ii) for every $\omega \in A$ and for infinitely many t's

$$\sup\{f(x);x \in P_t(\omega)\}\tilde{\mu}(P_t(\omega))/\mu(P_t(\omega)) > 1 + \varepsilon$$

or

$$\inf\{f(x);x \in P_t(\omega)\}\tilde{\mu}(P_t(\omega))/\mu(P_t(\omega) < 1 - \varepsilon.$$

Hence, for every $\omega \in A$ there are infinitely many t's and $B_t \subseteq P_t(\omega)$ s.t. (i) $\mu(B_t) > 0$ and (ii) either $f(x)/\mu(P_t(\omega)) > t + \varepsilon/2$ or $f(x)/\mu(P_t(\omega)) < 1 - \varepsilon/2$ for every $x \in B_t$. Therefore, either

$$\mu(B_t|P_t(\omega))\tilde{\mu}(P_t(\omega)) = [\mu(B_t)/\mu(P_t(\omega))]\tilde{\mu}(P_t(\omega)) < \tilde{\mu}(B_t)(1 - \varepsilon/2)$$

or

$$[\mu(B_t)/\mu(P_t(\omega))]\tilde{\mu}(P_t(\omega)) > \tilde{\mu}(B_t)(1 + \varepsilon/2).$$

Thus, either

$$\mu(B_t | P_t(\omega))/\tilde{\mu}(B_t | P_t(\omega)) > 1 + \varepsilon/2$$

or

$$\mu(B_t | P_t(\omega))/\tilde{\mu}(B_t | P_t(\omega)) < 1 - \varepsilon/2$$

for every $\omega \in A$ and for infinitely many t's. We conclude that $\tilde{\mu}$ does not $\{\mathcal{P}_t\}$-s.m.t. μ. //

<u>Corollary 2</u>: If $\tilde{\mu}$ $\{P_t\}$-s.m.t. μ for every information sequence, then the Radon Nikodym derivative $d\mu/d\tilde{\mu}$ is constant over atoms of $\tilde{\mu}$. Moreover, for every information sequence $\{\mathcal{P}_t\}$ there is time $t(\omega)$, s.t. $t > t(\omega)$ implies

$$\mu(\bullet | P_t(\omega)) = \tilde{\mu}(\bullet | P_t(\omega)).$$

<u>Corollary 3</u>: If $\tilde{\mu}$ $\{P_t\}$-s.m.t. μ for every information sequence, and $\tilde{\mu}$ is non-atomic. Then $\tilde{\mu} = \mu$.

References

Aumann, R. J. (1976), "Agreeing to Disagree," The Annals of Statistics, 4, 1236-1239.

Blackwell, D. and L. Dubins (1962), "Merging of Opinions with Increasing Information," Annals of Mathematical Statistics, 38, 882-886.

Breiman, L. (1968), Probability, Reading, Mass.: Addison-Wesley.

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

Halmos, P. R. (1950), Measure Theory, Springer-Verlag.

Kalai, E. and E. Lehrer (1993a), "Rational Learning Leads to Nash Equilibrium," Econometrica, forthcoming.

Kalai, E. and E. Lehrer (1993b), "Subjective Equilibrium in Repeated Games," Econometrica, forthcoming.

Kalai, E. and E. Lehrer (1990), "Merging Economic Forecasts," Northwestern University Discussion Paper.

Monderer, D. and D. Samet (1990), "Stochastic Common Learning," mimeo.

Schervish, M. and T. Seidenfeld (1990), "An Approach to Consensus and Certainty with Increasing Evidence," Journal of Statistical Planning and Inference, 25, 401-414.

Shiryayev, A. N. (1984), Probability, Springer-Verlag.