Divisive Conditioning: Further Results on Dilation*

Timothy Herron†‡
Department of Philosophy, Carnegie Mellon University

Teddy Seidenfeld
Departments of Philosophy and Statistics, Carnegie Mellon University

Larry Wasserman
Department of Statistics, Carnegie Mellon University

Conditioning can make imprecise probabilities uniformly more imprecise. We call this effect "dilation". In a previous paper (1993), Seidenfeld and Wasserman established some basic results about dilation. In this paper we further investigate dilation on several models. In particular, we consider conditions under which dilation persists under marginalization and we quantify the degree of dilation. We also show that dilation manifests itself asymptotically in certain robust Bayesian models and we characterize the rate at which dilation occurs.

1. Introduction. An important part of the Bayesian statistical lore is that increasing shared evidence leads (almost surely) to merging opinions, provided those opinions are not extremely discrepant to begin with. This is one way to rebut charges of excessive subjectivism in Bayesian epistemology: use interpersonal agreement to show that objectivity results even though probability is "personal." In his classic discussion of Bayesian inference, Savage (1972, §3.6 and 4.6) illustrates

*Received July 1996.

†Send reprint requests to the second or third author, Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890.

‡The first two authors were supported by NSF Grant SES-9208942. The third author was supported by NSF Grants DMS-9005858 and DMS-9357646 and NIH Grant RO1-CA54852-01.

Philosophy of Science, 64 (September 1997) pp. 411–444. 0031-8248/97/6403-0002$2.00

almost certain asymptotic consensus for finitely many investigators when the following assumptions hold:

(1) there are finitely many statistical hypotheses of interest;
(2) the investigators hold common likelihoods for these hypotheses given the shared data, and they agree that the data are i.i.d. (identically, independently distributed) given each statistical hypothesis; and
(3) their priors agree which of these statistical hypotheses has 0 prior probability.

Blackwell and Dubins (1962) reduce the three assumptions to one: interpersonal agreement about which (infinite) data sets have 0 prior probability. Stated technically, the (finitely many) agents' priors are required to be mutually absolutely continuous (m.a.c.). However, the result about merging of posteriors does not require either the data be conditionally i.i.d. or that there be only finitely many hypotheses.

What happens asymptotically, almost surely, is not always a useful guide to the short run. Seidenfeld and Wasserman (1993) address this question in the following way. Let M be a nonempty set of probability measures on a measurable space $(\Omega, \mathcal{A})$. The lower probability $\underline{P}$ and the upper probability $\overline{P}$ are defined by $\underline{P}(A) = \inf_{P\in M} P(A)$ and $\overline{P}(A) = \sup_{P\in M} P(A)$. If $\underline{P}(B) > 0$ define $\underline{P}(A|B) = \inf_{P\in M} P(A|B)$ and $\overline{P}(A|B) = \sup_{P\in M} P(A|B)$. A measurable partition $\mathcal{B}$ *weakly dilates* the event A if

$$\underline{P}(A|B) \leq \underline{P}(A) \leq \overline{P}(A) \leq \overline{P}(A|B) \text{ for all } B \in \mathcal{B}. \quad (1)$$

We say that $\mathcal{B}$ *dilates* the event A if (1) holds with at least one of the outer inequalities being strict for some $B \in \mathcal{B}$. $\mathcal{B}$ *strictly dilates* the event A if both outer inequalities in (1) are strict for all $B \in \mathcal{B}$.

Upper and lower probability theory is one alternative to strict Bayesian methodology. First, upper and lower probabilities provide a rigorous mathematical framework for studying sensitivity and robustness in classical and Bayesian inference (Berger 1984, 1985, 1990; Lavine 1991; Huber and Strassen 1973; Walley 1991; Wasserman and Kadane 1992). Second, they arise in group decision problems (Levi 1982; Seidenfeld, Kadane, and Schervish 1989). Third, they can be justified by an axiomatic approach to uncertainty that arises when the axioms of probability are weakened (Good 1952, Smith 1961, Kyburg 1961, 1974, Levi 1974, Seidenfeld, Schervish and Kadane 1995, Walley 1991). Fourth, sets of probabilities may result from incomplete or partial elicitation. Finally, there is some evidence that certain physical phenomena may be described by lower and upper probabilities (Fine 1988, Walley and Fine 1982).

If $\mathcal{B}$ dilates A then observing $B \in \mathcal{B}$ is certain to increase the indeterminacy of the observer's beliefs about A. Alternatively, M may represent the beliefs of a set of observers; each $P \in M$ corresponds to one observer. In this case, observing $B \in \mathcal{B}$ increases the disagreement in the group about A. In either case, if $\mathcal{B}$ dilates A then there is some question about whether the experiment consisting of observing $B \in \mathcal{B}$ is worthwhile.

Seidenfeld and Wasserman (1993) showed that dilation is easily induced. Indeed, only very special classes of probabilities are immune to dilation. This paper is a continuation of that investigation. We begin with some propositions about dilation for ε-contamination classes of probabilities. These serve as a springboard for §2–7, where we discuss parallel results for other classes of probabilities.

Let P be a probability measure, let $\mathcal{P}$ be the set of all probability measures on $\mathcal{A}$ and define the *ε-contamination neighborhood* of P by

$$M = \{(1 - \varepsilon)P + \varepsilon Q;\ Q \in \mathcal{P}\} \tag{2}$$

where $\varepsilon \in [0,1]$. This class is ubiquitous in frequentist and Bayesian robustness theory (Huber 1981; Berger 1984, 1990). It can be understood in several ways. It is a model of data contamination where, with probability $(1 - \varepsilon)$ the intended statistical model P generates the data and with probability ε any other probability law Q generates the data. Alternatively, as we note below, this model corresponds to a class of probabilities formed by fixing lower bounds on the probabilities of atomic events in the algebra $\mathcal{A}$. Thus, it might be created by eliciting a set of "experts" for their probability judgments about the atoms of $\mathcal{A}$ and forming the convex closure of the lower probability bounds from this set. Yet another example is found in Walley's (1996) "imprecise Dirichlet" model for inference from multinomial data.

It follows that for $A \neq \emptyset$ or Ω, $\underline{P}(A) = (1 - \varepsilon)P(A)$ and $\overline{P}(A) = (1 - \varepsilon)P(A) + \varepsilon$. Fix an event A such that $0 < \underline{P}(A) \leq \overline{P}(A) < 1$. Let $\mathcal{B}$ be a finite, nontrivial partition for A by which we mean that $0 < P(B \cap A) < P(A)$ and $0 < P(B \cap A^c) < P(A^c)$ for all $B \in \mathcal{B}$. Likewise, it is a simple calculation to verify that $\underline{P}(A|B) = (1 - \varepsilon)P(AB)/((1 - \varepsilon)P(B) + \varepsilon)$ and that $\overline{P}(A|B) = ((1 - \varepsilon)P(AB) + \varepsilon)/((1 - \varepsilon)P(B) + \varepsilon)$. As shown in Seidenfeld and Wasserman 1993,

PROPOSITION 1: *$\mathcal{B}$ dilates A if and only if*

$$-\varepsilon P(A^c)P(B^c) < P(AB) - P(A)P(B) < \varepsilon P(A)P(B^c) \text{ for all } B \in \mathcal{B}. \tag{3}$$

Note, in particular, that if A and B are nearly independent under P for all $B \in \mathcal{B}$, then dilation occurs. Define the *extent of dilation* by $\Delta(A, \mathcal{B}) = \min_{B \in \mathcal{B}}[(\overline{P}(A|B) - \overline{P}(A)) + (\underline{P}(A) - \underline{P}(A|B))]$.

PROPOSITION 2: *For the ε-contamination model we have*

$$\Delta(A, \mathcal{B}) = \min_{B} \frac{\varepsilon(1 - \varepsilon)P(B^c)}{\varepsilon + (1 - \varepsilon)P(B)}. \tag{4}$$

Note that $\Delta(A, \mathcal{B})$ does not depend on the event A. Further, this quantity is maximized when $\varepsilon = \sqrt{P(B_*)}/(\sqrt{P(B_*)} + 1)$ if B_* is an element of $\mathcal{B}$ that minimizes (4).

To find out whether there is some partition that dilates an event A, apparently one must consider all possible partitions. Say that M has the *binary dilation property* if, whenever there exists a partition that dilates an event A there also exists a binary partition that dilates A. When this property holds we need only examine all binary partitions to see if dilation occurs. A partition $\mathcal{C}$ is a *coarsening* of the partition $\mathcal{B}$ if each $C \in \mathcal{C}$ is a union of elements of $\mathcal{B}$. Say that M satisfies the *coarsening property* if, whenever a partition with three or more elements $\mathcal{B}$ dilates A, there is a coarsening $\mathcal{C}$ of $\mathcal{B}$ such that $\mathcal{C}$ dilates A and $\mathcal{C}$ has strictly smaller cardinality than $\mathcal{B}$. Note that the coarsening property implies the binary dilation property. When the coarsening property obtains for a model, it means that dilation in that model is not related to the complexity of the partition or experiment involved. Also, we have an interest in coarsening as it relates dilation to measures of association in the ever popular study of 2 by 2 contingency tables (see, for example, Corollary 2.12).

If P is a probability on $(\Omega, \mathcal{A})$ and $\mathcal{B}$ is a finite, measurable partition, then $P_{\mathcal{B}}$ denotes the marginal of P on the algebra generated by $\mathcal{B}$. Also, $M_{\mathcal{B}} = \{P_{\mathcal{B}}; P \in M\}$.

PROPOSITION 3: *The ε-contamination class possesses the coarsening property.*

When dilation does not occur, we might hope for the opposite, that some narrowing of the set of probabilities occurs. That is, it would be reassuring if there are some instances where the probabilities become more precise. Say that *$\mathcal{B}$ constricts A* if

$$\underline{P}(A) \leq \underline{P}(A|B) \leq \overline{P}(A|B) \leq \overline{P}(A) \text{ for all } B \in \mathcal{B} \tag{5}$$

with at least one of the outer inequalities being strict for some B. As it turns out, this never happens in the ε-contamination model.

PROPOSITION 4: *For any A, if $\varepsilon > 0$ then there is no $\mathcal{B}$ which constricts A.*

In light of the Blackwell-Dubins (1962) result, alluded to in the opening paragraph, it is relevant to inquire whether dilation is a phenomenon that is mitigated asymptotically. Suppose that Ω is the parameter space for a statistical model and consider using a class of priors M for the parameter. ε-contamination classes are popular for such purposes (see Berger 1984, 1985, 1990 and Berger and Berliner 1986). Specifically, suppose that $X_n \sim N(\theta, 1/n)$ and let M be an ε-contaminated class of priors for θ. Let $\underline{P}(A|X_n = x_n) = \inf_{P \in M} P(A|X_n = x_n)$, let $\hat{\theta}$ be the maximum likelihood estimator and let $A_n = [\hat{\theta} - a_n, \hat{\theta} + a_n]$ where $a_n > 0$. Thus A_n corresponds to the usual $1 - \alpha$ level confidence interval estimate for θ or, what amounts to the same thing, an interval estimate obtained by inverting on a family of unrejected classical point-null hypotheses from tests whose observed significance levels are bounded below at α. It is also the usual credible region—a region whose posterior probability is bounded below at $1 - \alpha$—except for an asymptotically negligible error which depends on the prior. We say that A_n *asymptotically dilates* if

$$\lim \underline{P}(A_n|X_n = x_n) = 0 \ a.s.$$

as $n \to \infty$. Otherwise, we say that A_n is asymptotically dilation immune.

PROPOSITION 5: *Suppose that $a_n = \{n^{-1}(C + k \log n)\}^{1/2}$ for some constants C and $k > 0$. If M is an ε-contaminated model then A_n is dilation immune if and only if $k \geq 1$.*

The proof of this proposition follows from results in Pericchi and Walley 1991. The implication is that the usual credible regions of length $O(1/\sqrt{n})$ are not robust in the sense that they are asymptotically dilated. Regions of length $O(\sqrt{\log(n)/n})$ are needed to stop the lower bound from tending to 0.

REMARK: Proposition 5 does not contradict the Blackwell-Dubins result for two reasons: the ε-contamination model has more than finitely many extreme points and the extreme points include point masses which are not m.a.c. Our results in §6 show when asymptotic dilation of these credible regions occurs within a m.a.c. version of the model. Schervish and Seidenfeld (1990) provide other results about the asymptotic merging of posteriors in the case of infinitely many Bayesian opinions.

It is interesting to observe that this latter result is related to the behavior of acceptance regions in Bayesian testing. Specifically, consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Following Jeffreys (1961, Ch. 5) we use a prior of the form $Q = (1/2)\delta_{\theta_0} + (1/2) P$ where δ_{θ_0} is a point mass at θ_0 and $P(\{\theta_0\}) = 0$. The Bayes factor BF is defined to be the ratio of posterior odds of H_0 versus H_1 to prior odds. It is easily shown that $BF = L(\theta_0) \ / \int L(\theta) \ P(d\theta)$ where $L(\cdot)$ is the likelihood function. Suppose we reject H_0 if $BF < c$ for some constant c. Although it is not standard practice to do so, we can consider inverting such a test to form a pseudo-confidence region. Let A be the set of all θ not rejected by such a test. It follows from the discussion in Jeffreys 1961 (p. 248) that $A = \hat{\theta} \pm O(\sqrt{\log(n)/n})$. We see that dilation immune credible regions for the ε-contaminated model have exactly the same form as regions from inverting a Jeffreys hypothesis test. For additional discussion of this idea, see §4 of Herron, Seidenfeld, and Wasserman 1994.

The remainder of this paper explores many of these issues for other classes. Sections 2–5 discuss existence and extent of dilation and the binary coarsening property for the following models: atomic, total-variation, Frechet classes and symmetric classes. In §6 we explore the asymptotics of dilation further. Concluding remarks are contained in §7. Appendix 1 contains additional material related to the total variation model studied in §3. Specifically, we give a partial proof for a more general version of Theorem 3.5. Appendix 2 shows how the coarsening results can be extended to infinite partitions. All proofs are contained in Appendix 3.

2. Atomic Models. Let Ω be a nonempty, finite set and let $\mathcal{A}$ be its power set. Say that M is an atomic lower and upper probability model (ALUP) if there exist functions $\beta, \gamma : \Omega \to I\!R$ such that

(i) $\beta(\omega) \leq P(\{\omega\}) \leq \gamma(\omega)$ for all $\omega \in \Omega$, where P is a probability distribution, implies that $P \in M$ and
(ii) for every $\omega \in \Omega$, $\sup_{P \in M} P(\{\omega\}) = \gamma(\omega)$ and $\inf_{P \in M} P(\{\omega\}) = \beta(\omega)$.

We say that M is generated (β, γ). This is similar to the density bounded classes studied by Lavine (1991) except that he does not require (ii). Though ALUP models generalize the class of ε-contamination models (as explained below), they are not as general as the class of upper and lower probability models on events in $\mathcal{A}$ developed by C. A. B. Smith's unconditional pignic odds (1961) or Kyburg's (1974) intervals of "epistemological" probabilities and, a fortiori, not as general as the complete class of convex sets of probability distributions on $\mathcal{A}$ used by Levi (1974, 1980). Nonetheless, we develop dilation results for ALUP models en route to a future study of convex sets of probabilities.

Consider a discrete $2 \times n$ atomically generated algebra, with $\{a_{ij} | 1 \leq j \leq n, i = 1 \text{ or } 2\}$ as its atoms, as follows: $\mathcal{P}_{\beta\gamma} = \{0 < \beta_{a_{ij}} \leq P(a_{ij}) \leq \gamma_{a_{ij}} < 1 | 1 \leq j \leq n$ and i $= 1$ or 2 and the γ's and β's are constants$\}$. The β's are the lower bounds for each of the atoms, and the γ's are the upper bounds. We need to make sure that the upper and lower bounds given for an ALUP model are effective; e.g., that the given upper bound for atom a_{11} is determined by $\gamma_{a_{11}}$ and not the lower bounds of the other atoms (β_{a12}, β_{a22}, . . .). Hence, according to (ii), we stipulate that the following must hold: For all $\beta_{a_{ij}}$ there exists a $P \in \mathcal{P}_{\beta\gamma}$ such that $P(a_{ij}) = \beta_{a_{ij}}$, and for all $\gamma_{a_{ij}}$ there exists a $P \in \mathcal{P}_{\beta\gamma}$ such that $P(a_{ij}) = \gamma_{a_{ij}}$.

For the 2 × 2 case where the space of all probabilities can be represented as a tetrahedron, the ALUP model's set of probabilities is, typically, an 8-sided polyhedron sitting in the tetrahedron with 4 hexagonal faces (the lower bounds) and 4 triangular faces (the upper bounds). The orientation of the 8-sided polyhedron is such that one hexagonal face and one triangular face are parallel to a given face of the tetrahedron, with the hexagonal face being closer to the tetrahedral face.

REMARK: The ALUP model reduces to the ε-contamination model when $\beta_{a_{ij}} = \varepsilon P(A_{ij})$ and $\gamma_{a_{ij}} = (1 - \varepsilon)P(A_{ij}) + \varepsilon$. Likewise, each ε-contamination model is ALUP by the same constraints.

2.1 Dilation Conditions for the ALUP Model. Define X to be the set of atoms $\{a_{ij} | 1 \leq j \leq n$ and $i = 1, 2\}$, and let $\mathcal{A}$ be the algebra over the atoms of X. Set $A = \cup_{j=1}^{n} a_{1j}$ and $b_j = a_{1j} \cup a_{2j}$ in this 2 × n atomic case. Also, define $[\gamma_E] = \{P \in \mathcal{P}_{\beta\gamma}; P(E) = \gamma_E\}$ and $[\beta_E] = \{P \in \mathcal{P}_{\beta\gamma}; P(E) = \beta_E\}$ where E is an event in $\mathcal{A}$ and γ_E and β_E are, respectively, the maximum and minimum probability values for the event *E,* for probabilities in $\mathcal{P}_{\beta\gamma}$. Note that $[\gamma_E]$ and $[\beta_E]$ are closed, convex sets since they are supporting for the ALUP model. Theorem 2.1 and Corollary 2.2 report some facts about extreme probability values for events and atoms they contain. Corollary 2.3 gives necessary and sufficient conditions for dilation in an ALUP model by giving separate conditions for dilation of the upper and lower probability values.

THEOREM 2.1: *If* $a_{ij} \in E \in \mathcal{A}$ *then* $[\gamma_E] \cap [\gamma_{a_{ij}}] \neq \emptyset$ *and* $[\beta_E] \cap [\beta_{a_{ij}}] \neq \emptyset$.

COROLLARY 2.2: *For all* $a_{ij} \neq a_{kl}$ *there exists* $P \in \mathcal{P}_{\beta\gamma}$ *such that* $P(a_{ij}) = \beta_{a_{ij}}$ *and* $P(a_{kl}) = \gamma_{a_{kl}}$.

COROLLARY 2.3: *(a)* $\overline{P}(A) < \overline{P}(A|b_j)$ *iff* $\dfrac{\gamma_{a_{1j}}}{\gamma_A} - \dfrac{\beta_{a_{2j}}}{\beta_{A^c}} > 0$.

(b) $\underline{P}(A) > \underline{P}(A|b_j)$ *iff* $\dfrac{\gamma_{a_{2j}}}{\gamma_{A^c}} - \dfrac{\beta_{a_{1j}}}{\beta_A} > 0$.

2.2 Coarsening in the ALUP Model. We keep the same 2 × n ALUP model that we had in the last subsection. The principal result of this subsection is the following theorem, which asserts that ALUP models have the coarsening property.

THEOREM 2.4: *Assume that for every* $j = 1, \ldots, n$ $n \geq 3$ *we have*

$\underline{P}(A|b_j) < \underline{P}(A) \leq \overline{P}(A) < \overline{P}(A|b_j)$; *in other words the partition* $\mathcal{B} = \{b_1, \ldots, b_n\}$ *strictly dilates A. Then there exists some coarsening* $\mathcal{B}' = \{b'_1, \ldots, b'_m\}$ *strictly dilating A.*

As with the rest of the results, the proof is in Appendix 3. The proof makes use of several Lemmas which are recorded here for the interested reader.

LEMMA 2.5: (Marginalization Property for ALUP Models) *Let* $\mathcal{P}$ be the ALUP model described above, and let E be an event in the algebra $\mathcal{A}$ of the model. Let $X_E = \{D \subset X - E\} \cup \{E\}$. *Thus, the algebra* $\mathcal{A}_E$ *is generated by the atoms of* X_E *which we define to be the coarsened subalgebra of* $\mathcal{A}$. Finally, define $\mathcal{P}_E$ *as the set of induced probabilities over the algebra* $\mathcal{A}_E$. *Then* $\mathcal{P}_E$ *is an ALUP model with constraints* $\beta'_a \leq P(a) \leq \gamma'_a$ *for all atoms* $a \in X_E$.

We need a few definitions for the next three Lemmas. If $\mathcal{G} = \{E_i\}$ is a family of subsets of $\{b_1, \ldots, b_n\}$, then we say that $\mathcal{G}$ *spans X* if $\cup_{\mathcal{G}} E_i = X$. We say that $\mathcal{G}$ *spans X with focus* b_m if, in addition, for all $E_i \neq E_k \in \mathcal{G}$ we have $E_i \cap E_k = \emptyset$ or $E_i \cap E_k = b_m$, and there exists $E_i \neq E_k \in \mathcal{G}$ such that $E_i \cap E_k = b_m$. Also, we call $\mathcal{G}$ a *tiling* of X if $\mathcal{G}$ spans X and for all $E_i \neq E_k \in \mathcal{G}$ we have $E_i \cap E_k = \emptyset$. Finally, we define an event E as an *upper (lower) dilator* for A if $\overline{P}(A) < \overline{P}(A|E)$ $(\underline{P}(A) > \underline{P}(A|E))$.

LEMMA 2.6: *Assume that the conditions of Theorem 2.4 hold. Let* $\mathcal{G}$ span X with focus b_k. *Then, (1)* $\mathcal{G}$ *contains an upper dilator for A. (2)* $\mathcal{G}$ *contains a lower dilator for A.*

Define $E^j = \{b_1, \ldots, b_{j-1}, b_{j+1}, \ldots, b_n\}$ for $j = 1, \ldots, n$.

LEMMA 2.7: *Assume that the antecedent of Theorem 2.4 holds. Then,*

(1) At least one E^j *is an upper dilator for A.*

(2) At least one E^j *is a lower dilator for A.*

LEMMA 2.8: *Assume the antecedent of Theorem 2.4 holds but that the conclusion of the theorem is false. If* $\mathcal{G} = \{E_{ij}\}$ *is a tiling of X, where* $E_{ij} = \{b_i, b_j\}$ *so that n is an even number, then*

(1) $\mathcal{G}$ contains at least one upper dilator for A.

(2) $\mathcal{G}$ contains at least one lower dilator for A.

We prove Theorem 2.4 by playing the following formal game: Define $\mathcal{G}_i = \{E_{ij} | 1 \leq i \neq j \leq n\}$, and $\mathcal{G} = \cup_i \mathcal{G}_i$. We can think of $\mathcal{G}$ as forming a two dimensional table with rows being the $\mathcal{G}_i$, $i = 1, \ldots, n$. We label each $E_{ij} \in \mathcal{G}$ as one of U, L, N, B depending upon whether the E_{ij} in question is an upper dilator but not a lower dilator, a lower dilator but not an upper dilator, neither an upper or lower dilator, or both an upper and a lower dilator, respectively. Notice that if some E_{ij} in our $\mathcal{G}$ is labeled B, we can construct the coarsening $\{b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_{j-1}, b_{j+1}, \ldots, b_n, E_{ij}\}$ which dilates A. Thus, Theorem 2.4

would be proved. Hence, we only have left to show that we cannot construct an ALUP model such that $\mathcal{G}$ has no E_{ij}'s in it with the label B. We shall do this by using Lemmas 2.6 and 2.8, which, translated into the game we are playing appear as:

LEMMAS 2.6 and 2.8:

If $C = \{E_{ij}\}$ is a set of pairs $E_{ij} = \{b_i, b_j\}$ that spans X with focus b_k or as a tiling of X, then either

1) Some $E_{ij} \in C$ is labeled B.

2) One $E_{ij} \in C$ is labeled U and another $E_{kl} \in C$ is labeled L.

We will assume that case 1) in the Lemmas never happens. Thus, each focused span of X or tiling of X places a restriction on the possible labellings that $\mathcal{G}$ can contain. In fact, our task is made easier by the following Lemma.

LEMMA 2.9: *If there is a labeling of $\mathcal{G}$ possible under the restrictions of Lemmas 2.6 and 2.8 when we have labels U, L, and N available, then there is a labeling of $\mathcal{G}$ possible using only the labels U and L obeying the restrictions of 2.6 and 2.8.*

COROLLARY 2.10: *Under the conditions of Theorem 2.4, there is a binary subpartition that strictly dilates A.*

PROOF. Iterate the theorem until $m = 2$. □

2.3 Extent of Dilation in ALUP Models. If we have a probability measure P on our partition $\{b_1, \ldots, b_n\}$, then we will define $\delta_P(A, b_j) = P(A \cap b_j)\, P(A^c \cap b_j^c) - P(A^c \cap b_j)\, P(A \cap b_j^c)$. We can derive the following Lemma using this notation, and produce an interesting connection between the extent of dilation and covariance in a corollary.

LEMMA 2.11: $\Delta(A, \mathcal{B}) = \min_j$

$$\left\{(\beta_A - \gamma_A) + \frac{\gamma_{a1j}\gamma_{a2j} - \beta_{a1j}\beta_{a2j}}{(\beta_{a1j} + \gamma_{a2j})\,(\gamma_{a1j} + \beta_{a2j})}\right\} = \min_j \left\{\frac{\delta_{P_{1j}}(A,b_j)}{P_{1j}(b_j)} - \frac{\delta_{P_{2j}}(A,b_j)}{P_{2j}(b_j)}\right\}$$

where $P_{1j} \in [\gamma_A] \cap [\gamma_{a1j}] \cap [\beta_{a2j}]$ and where $P_{2j} \in [\beta_A] \cap [\beta_{a1j}] \cap [\gamma_{a2j}]$.

COROLLARY 2.12: *If P is a probability on an ALUP model, then $\delta_P(A, b_j) = covariance(\chi_A, \chi_{b_j})$, where χ_E is the characteristic function on event E. Hence,* $\Delta(A, \mathcal{B}) = \min_j \left\{\frac{cov(\chi_A, \chi_{b_j})}{P_{1j}(b_j)} - \frac{cov(\chi_A, \chi_{b_j})}{P_{2j}(b_j)}\right\}$.

3. The Total Variation Model. The Blackwell-Dubins (1962) result about the almost-sure asymptotic merging of two posterior distributions, based on increasing shared evidence, uses the total variation metric, $\rho(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$, to measure the difference between two distributions P and Q. This provides a conservative measure of agreement since when a sequence of probability distributions converges in total variation, it has uniform (rather than pointwise) convergence.

First we discuss characterizations of dilation for the total variation set of probabilities. This is a model $M = \{Q;\ \rho(P, Q) \leq \varepsilon\}$ such that $\rho(P, Q) = \sup_A |P(A) - Q(A)|$ where ε is a fixed real between 0 and 1 and P is a fixed probability distribution. Thus M consists of a focal distribution P and all Q's within a total variation distance ε of Q. We also assume that P is internal meaning that P assigns positive mass to each atom in $\mathcal{A}$.

We obtain $\underline{P}(A) = \max\{P(A) - \varepsilon, 0\}$ and

$$\underline{P}(A|B) = \frac{\max\{P(AB) - \varepsilon, 0\}}{\max\{P(AB) - \varepsilon, 0\} + \min\{P(A^cB) + \varepsilon, 1\}}. \tag{6}$$

We also can derive analogous equations for the upper probabilities. In what follows we define

$d_P(A,B) = P(AB) - P(A)P(B)$ and
$S_P(A, B) = P(AB)/(P(A)P(B))$ if $P(A)P(B) > 0$ and
$S_P(A, B) = 1$ if $P(A)P(B) = 0$.

3.1 Characterizations of Dilation for Total Variation Sets. In Seidenfeld and Wasserman 1993 the following Lemma was proved:

LEMMA 3.1: *$\{B, B^c\}$ dilates A iff the following two conditions hold:*

(*i*) $\varepsilon > \max\{-d_P(A, B)/P(B^c), d_P(A, B)/P(B)\}$,

and

(*ii*) *if $P(AB^c) > \varepsilon$ then we must have that $\varepsilon > -d_P(A, B)/P(B)$, and if $P(AB) > \varepsilon$ then we must have that $\varepsilon > d_P(A, B)/P(B^c)$.*

Note that the Lemma only gives a characterization of dilation, *not* strict dilation. Also, we see that the Lemma breaks up the condition for dilation into four cases depending upon whether a probability in the set M can take the value zero on the events AB or AB^c.

The following Lemma characterizes when the total variation set of probabilities M forces B to strictly dilate the event A. Lemma 3.2 is needed for the coarsening result which follows.

LEMMA 3.2: *B strictly dilates A if and only if: (4 cases)*

1. $\varepsilon < \min\{P(A), P(A^c)\}$ *when* $P(AB), P(A^cB) < \varepsilon$
2. $P(A^c)(1 - S_P(A^cB^c)) < \varepsilon < P(A)$ *when* $P(AB) < \varepsilon \leq P(A^cB)$
3. $P(A)(1 - S_P(A, B^c)) < \varepsilon < P(A^c)$ *when* $P(A^cB) < \varepsilon \leq P(AB)$
4. $\varepsilon > \max\{P(A^c)(1 - S_P(A^c, B^c)), P(A)(1 - S_P(A, B^c))\}$ *when* $\varepsilon \leq P(AB), P(A^cB)$.

3.2 Coarsening of Dilation in Total Variation Sets. The following two

Lemmas are used in the coarsening theorem, and they can both be proven in a straightforward manner.

LEMMA 3.3: *Fix a probability distribution P over an algebra* $\mathcal{C}(\Omega)$, *and fix a real* ε *between 0 and 1. Let M be a total variation set of probabilities based on P and* ε. *If* $\mathcal{A}$ is a subalgebra of $\mathcal{C}(\Omega)$, *then defining* $M_{\mathcal{A}} = \{Q_{\mathcal{A}};\ Q \in M\}$ *we get* $M_{\mathcal{A}} = \{Q;\ \rho_{\mathcal{A}}(Q, P_{\mathcal{A}}) \leq \varepsilon\}$ *where* $P_{\mathcal{A}}$ *is the restriction of P to algebra* $\mathcal{A}$ and $\rho_{\mathcal{A}}$ *is the total variation distance on the algebra* $\mathcal{A}$.

LEMMA 3.4: *For a partition* $\{B_1, \ldots, B_n\}$ *of B we obtain*

1) $\sum_{i=1}^{n} S_P(A, B_i)P(B_i) = 1$ *if B is the entire probability space, and*

2) $S_P(A,B) = \sum_{i=1}^{n} \left(\frac{P(B_i)}{P(B)}\right) S_P(A,B_i)$ *for any B.*

The desired coarsening result for the total variation set of probabilities is now presented. The theorem is proved assuming that M is sufficiently far from the edges of the simplex. Other cases are discussed in Appendix 1.

THEOREM 3.5: *Let M be a total variation neighborhood of probability distributions centered on P where* ε *is fixed. Suppose that* $\pi_n = \{C_1, \ldots, C_n\}$ *is a partition of* Ω *with n elements which strictly dilates the event A* $(n \geq 3)$. *For all* $i = 1, \ldots, n$ *assume that* $P(AC_i)$ *and* $P(A^cC_i) \geq \varepsilon$. *Then, there exists a binary subpartition* $\{B, B^c\}$ *which dilates A also.*

3.3 Extent of Dilation for the Total Variation Set of Probabilities. By straightforward derivations using calculations performed in Lemma 3.2, we can prove the following proposition:

PROPOSITION 3.6: *The extent of dilation for the total variation model breaks into four cases:*

a. *If* $P(AB_i), P(A^cB_i) < \varepsilon$, *then* $\Delta(A, \{B_i\}) = 1 - \min\{1, P(A) + \varepsilon\} + \max\{0, P(A) - \varepsilon\}$.
b. *If* $P(AB_i) < \varepsilon \leq P(A^cB_i)$ *then,* $\Delta(A, \{B_i\}) = (P(AB_i) + \varepsilon)/P(B_i) - (P(A) + \varepsilon - \max\{0, P(A) - \varepsilon\})$.
c. *If* $P(A^cB_i) < \varepsilon \leq P(AB_i)$ *then,* $\Delta(A, \{B_i\}) = (P(A^cB_i) + \varepsilon)/P(B_i) - (P(A^c) + \varepsilon - \max\{0, P(A^c) - \varepsilon\})$.
d. *If* $\varepsilon \leq P(AB_i), P(A^cB_i)$ *then,* $\Delta(A, \{B_i\}) = 1 - P(B_i)$.

COROLLARY 3.7: $\Delta(A, \{B_i\}) \geq 0$ *in a. and d. of Proposition 3.6. The condition for* $\Delta(A, \{B_i\}) \geq 0$ *in cases b and c may or may not be equivalent to the condition for* B_i *dilating A.*

As an example of a case of negative extent of dilation from case c. we take $\varepsilon = .2$,

$P(AB) = .69$, $P(A^cB) = .01$, $P(AB^c) = .1$, and $P(A^cB^c) = .2$. Then, $\underline{P}(A) = .59$, $\underline{P}(A|B) = .70$, $\overline{P}(A) = .99$, and $\overline{P}(A|B) = 1.0$ (rounded to two decimal places).

4. Frechet Classes. If X and Y are random variables with distribution P_X and P_Y respectively, then the *Frechet class* is defined to be the set of all joint probability measures for X and Y with the given marginals P_X and P_Y (Rachev 1985). First we consider only the simplest of Frechet models, the set of probabilities of a finite partition with fixed marginals.

Thus, we have $\Omega = \{AB_1, A^cB_1, AB_2, A^cB_2, \ldots, A^cB_n\}$, and for some fixed r, $c_i \in [0,1]$ i $= 1, \ldots,$ n such that $\Sigma_{i=1}^n c_i = 1$ we let $\mathcal{P} = \{P|P(A) = r \text{ and } P(B_i) = c_i, i = 1, \ldots, n\}$.

4.1 Dilation Conditions for the Fixed Marginals Model. Strict dilation occurs often in this model as evidenced by Theorem 4.2.

LEMMA 4.1:

$$\underline{P}(A|B_i) = \frac{max\{0, r + c_i - 1\}}{c_i}, \ \overline{P}(A|B_i) = \frac{min\{r, c_i\}}{c_i}. \tag{7}$$

THEOREM 4.2: *The partition $\mathcal{B} = \{B_1, \ldots, B_n\}$ strictly dilates A if and only if r and all of the c_i have values other than 0 or 1.*

4.2 Coarsening of Dilation on the Fixed Marginals Model.

THEOREM 4.3 *If a partition B strictly dilates A, then any nontrivial coarsened subpartition B′ of B also strictly dilates A.*

Regarding a more general Frechet class, we assert the following result: Let F_X and F_Y each be a c.d.f. on the interval (0,1). Following Dall'Aglio 1972, define the two dimensional Frechet class on F_X and F_Y by: $\Gamma(F_X, F_Y)$ is the set of joint distribution functions for X and Y with marginals F_X and F_Y for X and Y respectively. Let A be the event $(a,b) \times (0,1)$, where $0 < a < b < 1$, and let B be the event $(0,1) \times (c,d)$, where $0 < c < d < 1$. Assume that neither A nor B is empty of probability mass or contains all the probability mass. Then:

PROPOSITION 4.4: $\mathcal{B}$ dilates A.

As with the fixed margins model, we think that coarsenings to finite nontrivial subpartitions will preserve dilation.

4.3. Extent of Dilation for the Fixed Marginals Model. Simple calculations based on Lemma 4.1 yield:

PROPOSITION 4.5: $\Delta(A, \mathcal{B}) = \min_i \dfrac{\min\{r, 1-r, c_i, 1-c_i\}}{c_i}$.

COROLLARY 4.6: $\Delta(A, \mathcal{B}) > 0$ *unless some* r, $c_i \in \{0,1\}$.

5. Symmetric Classes. For this section $\Omega = [0,1]$, $\mathcal{A}$ is the *class of Borel subsets of* Ω *and* μ *is Lebesgue measure. We are concerned with neighborhoods of* μ. Many classes of probabilities are neighborhoods of some probability measure and, by an appropriate transformation, can be considered neighborhoods of μ. All ordinary neighborhoods of μ have some common structure (Wasserman and Kadane 1992) which we now describe.

Two density functions f and g are equimeasurable, denoted by $f \sim g$, if $\mu(f > t) = \mu(g > t)$ for all t, where $(f > t) \equiv \{(\omega \in \Omega; f(\omega) > t\}$. Given f, there is a unique, nonincreasing right continuous function f^* and a unique, nondecreasing right continuous function f_* such that $f \sim f^*$ and $f \sim f_*$. These are called, respectively, the *decreasing and increasing* rearrangements of f. Let $\Lambda(f) = \{g; g \sim f\}$ be the orbit of f. Let $u(\omega) = 1$ for all ω. We call m a *symmetric neighborhood* of u if $f \in m$, $g \sim f$ implies that $g \in m$. (Formally, each density should be replaced by an equivalence class of densities that represent the same probability measure; we shall not make such a distinction here.) m is a *density ratio class* (DeRobertis and Hartigan 1981) if there is a $k \geq 1$ such that

$$m = \left\{g; \frac{\text{ess sup } g}{\text{ess inf } g} \leq k\right\}. \tag{8}$$

(Recall the ess sup$f(\omega)$ is the supremum of all real numbers a such that $\mu(\{\omega; f(\omega) > a\}) > 0$ and ess inf is defined similarly.) In Seidenfeld and Wasserman 1993 it was shown that a symmetric neighborhood is immune to dilation if and only if its weak convex closure is a density ratio class. This does not tell us, however, for a given A whether there is a partition that dilates A. Now we take a closer look at these neighborhoods.

5.1 Conditions for Dilation and the Binary Dilation Property. Let $\overline{P}(A) = \sup_{f \in m} \int_A f(\omega)\mu(d\omega)$ and $\underline{P}(A) = \inf_{f \in m} \int_A f(\omega)\mu(d\omega)$. Let $\langle A\rangle^0 = \{f; \int_A f(\omega)\mu(d\omega) = \overline{P}(A)\}$ and $\langle A\rangle_0 = \{f; \int_A f(\omega)\mu(d\omega) = \underline{P}(A)\}$. It can be shown that $\langle A\rangle^0$ and $\langle A\rangle_0$ are nonempty. We begin with a Lemma that will be useful for further calculations; the proof involves straightforward algebra and is omitted.

LEMMA 5.1.1: *Let* f *be a density and consider real numbers* $0 = a_0 < a_1 < a_2 < a_3 < a_4 = 1$. *Let* $F(\omega) = \int_0^\omega f_*(t)\mu(dt)$ *and suppose that* $F(a_i) - F(a_{i-1}) > 0$, $i = 1,2,3,4$. *Then,* $F(a_1)/(F(a_1) + F(a_4) - F(a_3) \leq$

$F(a_2)$ *if and only if* $(F(a_4) - F(a_3))/(F(a_1) - F(a_0)) \geq (F(a_3) - F(a_2))/(F(a_2) - F(a_1))$.

Let $A \in \mathcal{A}$ be such that $0 < \mu(A) < 1$. A partition $\mathcal{B}$ is nontrivial if $0 < \mu(A \cap B) < \mu(A)$ and $0 < \mu(A^c \cap B) < \mu(A^c)$ for all $B \in \mathcal{B}$. We say that $\mathcal{B}$ is *balanced* if

$$\frac{\mu(A \cap B)}{\mu(A)} = \frac{\mu(A^c \cap B)}{\mu(A^c)} \text{ for all } B \in \mathcal{B}.$$

We say that $\mathcal{B}$ is *very balanced* if $\mu(A \cap B) = \mu(A \cap B')$ and $\mu(A^c \cap B) = \mu(A^c \cap B')$ for all $B, B' \in \mathcal{B}$.

LEMMA 5.1.2: *If $\mathcal{B}$ is balanced then $\mathcal{B}$ weakly dilates A.*

LEMMA 5.1.3: *There exists a binary partition that weakly dilates A.*

For every A and every f define

$$\phi(A,f) = \frac{\text{ess sup}_A f}{\text{ess inf}_A f}\frac{\text{ess sup}_{A^c} f}{\text{ess inf}_{A^c} f}. \tag{9}$$

Now define $\phi_0(A) = \sup_{f\in\langle A\rangle_0} \phi(A,f)$ and $\phi^0(A) = \sup_{f\in\langle A\rangle^0} \phi(A,f)$. Note that $\phi_0(A)\phi^0(A) \geq 1$.

LEMMA 5.1.4: *If $\phi_0(A)\phi^0(A) > 1$ then A has a strict, binary, dilator.*

The converse of Lemma 5.1.4 is not true in general. To see this, consider the following example. Let $f(\omega) = 1/2$ on $A = [0, 1/2]$ and $f(\omega) = 3/2$ on A^c. Let $B = [0, 1/4] \cup [1/2, 3/4]$ and define $g(\omega) = 1/16$ on $A \cap B$, $g(\omega) = 3/16 + \varepsilon$ on $A \cap B^c$ and $g(\omega) = 3/8 - \varepsilon/2$ on A^c where $\varepsilon < 3/8$. Let m consist of all rearrangements of f and g. Then $\langle A\rangle_0 = \{f\}$ and $\langle A\rangle^0 = \{f^*\}$ so that $\phi_0(A)\phi^0(A) = 1$. But $\underline{P}(A|B) = \underline{P}(A|B^c) = \{7 - 8\varepsilon\}^{-1} < \underline{P}(A) = 1/4$ and $\overline{P}(A|B) = \overline{P}(A|B^c) = (6 - 8\varepsilon)/(7 - 8\varepsilon) > \overline{P}(A) = 3/4$ so that strict dilation occurs.

There is, however, a converse under some stronger conditions. We say that $\overline{P}$ is 2-alternating if $\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B) - \overline{P}(A \cap B)$. We say that m is I-closed (I for indicator functions) if $P(A) \leq \overline{P}(A)$ for all $A \in \mathcal{A}$ implies that $dP/d\mu \in M$.

LEMMA 5.1.5: *Suppose that $\overline{P}$ is 2-alternating and m is I-closed. Then A has a strict dilator if and only if $\phi_0(A)\phi^0(A) > 1$.*

In addition, we have the following fact about these classes.

LEMMA 5.1.6: *Suppose that $\overline{P}$ is 2-alternating and that m is I-closed. Then constriction never occurs.*

5.2. Extent of Dilation. We shall assume that

$$0 < c = \inf_{f\in m} \text{ess inf} f \leq \sup_{f\in m} \text{ess sup} f < C < \infty$$

for some c and C. Define $f_\square = f_*(0)\{f_*(0) + f_*(1^-)\}^{-1}$ and $f^\square =$

$f^*(0)\{f^*(0) + f^*(1^-)\}^{-1} = f_*(1^-)\{f_*(1^-) + f_*(0)\}^{-1}$. Let $m_\square = \inf_{f\in m} f_\square$ and $m^\square = \sup_{f\in m} f^\square$. It follows that $0 < m_\square < 1/2 < m^\square < 1$.

LEMMA 5.2.1: Let Γ be the set of all nontrivial partitions. Then,

$$\inf_{\mathcal{B}\in\Gamma}\max_{B\in\mathcal{B}} \underline{P}(A|B) = m_\square$$

and

$$\sup_{\mathcal{B}\in\Gamma}\min_{B\in\mathcal{B}} \underline{P}(A|B) = m^\square.$$

A measure of the extent of lower dilation is $\sup_{\mathcal{B}\in\Gamma} \min_{B\in\mathcal{B}}[\underline{P}(A) - \underline{P}(A|B)] = \underline{P}(A) - \inf_{\mathcal{B}\in\Gamma} \max_{B\in\mathcal{B}} \underline{P}(A|B) = \underline{P}(A) - m_\square$. Similarly, a measure of the extent of upper dilation is $m^\square - \overline{P}(A)$.

6. Remarks on Asymptotic Dilation. As noted in the Introduction, when a set of priors is used for Bayesian inference, the usual credible regions of size $O(n^{-1/2})$ have the property that the lower posterior probability tends to 0. This may be viewed as a type of asymptotic dilation. We observed that regions of the form $X_n \pm a_n$ with $a_n = \{n^{-1}(C + k \log n)\}^{1/2}$ did not dilate if $k \geq 1$. On the other hand, the usual regions correspond to $k = 0$. This raises the following question: is it possible to have sets of priors where the dilation immune regions are in between these two extremes ($k = 0$ and $k = 1$)? The answer is yes. The rest of this section explores some intermediate cases.

Let $\Theta = (0, 1)$; any bounded, open subset of the real line will do. Let μ be Lebesgue measure which we regard as the base prior. Let $M = (1 - \varepsilon)\mu + \varepsilon\mathcal{S}$ where $0 < \varepsilon < 1$ and $\mathcal{S}$ is a symmetric class. Taking $\mathcal{S}$ to be all point masses gives the usual contamination neighborhood. Taking $\mathcal{S} = \{\mu\}$ gives a singleton set corresponding to the often used flat prior.

Let $X_n|\theta \sim N(\theta, 1/n)$ and let $A_n = X_n \pm a_n$ where $a_n > 0$ and $a_n = o(1)$. (Extensions to non-normal likelihoods are straightforward as long as we work on a bounded region and as long as the usual conditions are assumed to achieve asymptotic normality.) *Asymptotic dilation* means that

$$\lim_n \underline{P}(A_n|X_n = x_n) = 0 \text{ almost surely.} \tag{10}$$

The next two Lemmas pin down two extreme cases.

LEMMA 6.1: *If* $a_n = o(n^{-1/2})$ *then for any symmetric* $\mathcal{S}$, A_n *asymptotically dilates.*

The next Lemma is an easy consequence of the results in Walley and Pericchi 1991 and the proof is omitted.

LEMMA 6.2: *If* $a_n = \{n^{-1}(C + \log n)\}^{1/2}$. *Then, for any symmetric* $\mathcal{S}$ A_n *is asymptotically dilation immune.*

Now we construct classes with asymptotic behavior in between these two extremes. In light of the previous two Lemmas we assume that

$$\lim \sqrt{n}a_n > 0. \tag{11}$$

For $\alpha \in (0, 1)$ let $p_\alpha(\theta) = (1 - \alpha)\theta^{-\alpha}$. Let $\mathcal{S}_\alpha$ be all rearrangements of p_α. Now we establish the rate of a_n needed to avoid dilation for given α.

To begin, we first establish the probability measure that minimizes $P(A_n|X_n = x_n)$.

LEMMA 6.3: Without loss of generality, assume that $x_n \geq 1/2$. Also assume that n is large enough so that $x_n + a_n < 1$. Then, $\overline{P}(A_n|X_n = x_n) = J(A_n|X_n = x_n)$ *where*

$$\frac{dJ}{d\mu}(\theta) = (1 - \varepsilon) + \varepsilon r_n(\theta)$$

and the density r_n *is defined by*

$$(1 - \alpha)^{-1}r_n(\theta) = \begin{cases} (1 - \omega - 2a_n)^{-\alpha} & \text{if } \omega < 2x_n - 1 \\ (2(x - \omega - a_n))^{-\alpha} & \text{if } 2x_n - 1 \leq \omega < x_n - a_n \\ (1 - 2|x - \omega|)^{-\alpha} & \text{if } x_n - a_n \leq \omega < x_n + a_n \\ (2(\omega - x_n - a_n))^{-\alpha} & \text{if } x_n + a_n \leq \omega \leq 1. \end{cases}$$

LEMMA 6.4: *Assuming (11), asymptotic dilation occurs if and only if*

$$\lim \int L_n(\theta)r_n(\theta)d\theta < \infty \text{ a.s.} \tag{12}$$

where,

$$L_n(\theta) \propto \sqrt{n} \exp\{-n(\theta - x_n)^2/2\}$$

is the likelihood function.

LEMMA 6.5: *If*

$$\inf_{k>0} \lim \frac{a_n^2 + ka_n n^{-1/2}}{\log n/n} < \alpha \tag{13}$$

then asymptotic dilation occurs.

THEOREM 6.6: *Let* $a_n = \{n^{-1}(C + d \log n)\}^{1/2}$ *for constants* $C >$ 0 *and d. Asymptotic dilation occurs if and only if* $d < \alpha$.

REMARK: The symmetry of $\mathcal{S}$ is not crucial. What matters is the rate at which the densities in the class become unbounded. In particular, if $\mathcal{S}$ is a set of uniformly-bounded densities then $a_n = C/\sqrt{n}$ will yield a dilation immune region.

7. Concluding Remarks.

7.1 Summary. Beginning in §1 with the ε-contaminated model, $M = (1 - \varepsilon)P + \varepsilon Q$ where P is a focal probability and Q is an arbitrary probability, we emphasized the following four aspects of dilation:

(1) The quantity $B \in \mathcal{B}$ dilates event A if and only if A and B are nearly independent under the focal distribution P.
(2) The extent to which B dilates A depends only on ε and $P(B)$. Parenthetically, the antithesis of dilation, called "constriction," does not obtain for this model.
(3) If $B \in \mathcal{B}$ dilates A, then some binary partition obtained by coarsening B also dilates A.
(4) With an ε-contaminated class of priors, the asymptotic dilation for likelihood-based interval estimates of a normal mean based on i.i.d. data is related to Jeffreys' rule for converting (Bayesian) posterior probabilities from hypothesis testing into (Classical) significance levels.

In §2 we explored the so-called ALUP generalization of the ε-contamination model. An ε-contamination model is recovered by fixing the lower probabilities for each atom of the algebra $\mathcal{A}$. The ALUP model results from specifying both lower and upper probabilities for each atom. We provide parallel dilation results for aspects (1)–(3) in ALUP models, including characterization of the extent of dilation in terms of covariance.

In §3 we examined the Total Variation model, again, for the purpose of producing parallels to (1)–(3). The "coarsening" result, (3), was established for what we call the "uniform condition" case. (We conjecture that the result obtains for the remaining "nonuniform" case too.)

In §4 we considered the Frechet classes, defined by fixing the marginal distributions for a pair of random variables. Counterparts of (1)–(3) were obtained straightforwardly for the fixed margins models on a finite algebra. In particular, regarding (3), strict dilation is preserved under all nontrivial coarsenings.

In §5 we shifted attention to models that are symmetric neighborhoods of a distribution on Borel subsets of [0,1]. (Assuming that distribution is dominated by Lebesgue measure, then, without loss of generality we are able to work with neighborhoods of the uniform distribution.) Again, we provide counterparts to results (1)–(3).

In §6, we attended to problems involving asymptotic dilation for the usual credible regions of a parameter, based on a variant of the ε-contamination model for the priors. We showed how the rate at which the interval estimates shrink determines, for a given model, whether or not asymptotic dilation occurs. In a dual form, given a rate for shrinking credible intervals (as a function of sample size), we characterized which models for the priors will experience asymptotic dilation.

7.2 Open questions. Several related issues about dilation strike us as worthy of continued investigation. First, the (binary) coarsening results are obtained by rather different methods, depending upon the model. Is there a unified approach to solving this problem? We do not know even if there is a model where the coarsening result fails! In §5 we saw that the 2-alternating condition played a role, as is common in the theory of upper and lower probabilities generally. Is there a useful explanation for this connection?

Last, the asymptotic results in section 6 suggest that robust credible regions in Bayesian inference will shrink at a slower than usual rate. Except in the works of Pericchi and Walley (1991) and Jeffreys (1961, Ch. 5)—based on our inversion on Jeffreys's hypothesis tests—we have not seen any discussion of this phenomenon before. In light of dilation, we think that the asymptotics of robust Bayesian inference deserve further scrutiny.

APPENDIXES

Appendix 1 discusses some technical issues left over from §3.2. Appendix 2 deals with extensions of §2, 3, and 4 for countable partitions.

Appendix 1
Further Results on Coarsening of Dilation in Total Variation Sets

In §3.2 we proved the coarsening result for total variation sets under restrictive conditions. Some other cases are presented here.

As before, we assume that we are given a partition $\pi_n = \{C_1, C_2, \ldots, C_n\}$ which dilates the event A such that no coarsening $\{B_1, B_2, \ldots, B_m\}$ of π_n dilates A.

CASE 1: For some $i, j \in \{1, \ldots, n\}$ such that i is not equal to j we have $P(A(C_i \cup C_j)) < \varepsilon$ and $P(A^c(C_i \cup C_j)) < \varepsilon$.

The inequalities imply that $P(AC_i)$, $P(AC_j)$, $P(A^cC_i)$ and $P(A^cC_j)$ are all $< \varepsilon$. Hence, since π_n strictly dilates A it must be that $P(A)$, $P(A^c) \geq \varepsilon$ by Lemma 3.2. But the assumptions of case 1 and the previous equations, by using Lemma 3.2 again, imply that the subpartition $\pi_{ij} = \{C_1, C_2, \ldots, C_{i-1}, C_{i+1}, \ldots, C_{j-1}, C_{j+1}, \ldots, C_n, C_i \cup C_j\}$ also dilates A $-$ a contradiction.

CASE 2: For all $i, j = 1, \ldots, n$ we have $P(AC_i)$, $P(AC_j)$, $P(A^cC_i)$ and $P(A^cC_j) \geq \varepsilon$.

This case was proved in §3.2.

CASE 3: For all $i, j \in \{1, \ldots, n\}$ where $i \neq j$ we have $P(AC_i)$, $P(A^cC_i) < \varepsilon$ and $P(A(C_i \cup C_j))$, $P(A^c(C_i \cup C_j)) \geq \varepsilon$.

It is not the case that any binary coarsening, π_{ij}, will strictly dilate A under the assumption of Case 3. For a counterexample we take the partition $\{C_1, C_2, C_3\}$ such that $P(AC_1) = 0.05$ and the five other atoms have probability 0.19. One can check that while π_3 strictly dilates A, π_{23} does not strictly dilate A (check by using Lemma 3.2).

However, we now show that every partition π_n strictly dilating A has some strictly dilating coarsening using a very different technique than the combinatorial approach used for Case 2 of this proof. The main idea is to use our special assumptions of Case 3 to show that the probability has to be spread out very evenly over the atoms. Then we find that such a relatively uniform distribution satisfies the equations in Lemma 3.2 (else we get a contradiction).

Using the assumptions $P(A(C_i \cup C_j)) \geq \varepsilon$ and $P(A^c(C_i \cup C_j)) \geq \varepsilon$ we see that case 4 of Lemma 3.2 applies. Hence, we can straightforwardly calculate the following strict dilation equivalences:

$$\text{If } S_P(A, C_i \cup C_j) > 1 \textit{ then } C_i \cup C_j \text{ strictly dilates } A \textit{ iff}$$

$$\varepsilon - P(A))P(C_i \cup C_j) + P((C_i \cup C_j)A) < \varepsilon. \tag{14}$$

$$\textit{If } S_P(A, C_i \cup C_j) < 1 \textit{ then } C_i \cup C_j \textit{ strictly dilates } A \textit{ iff}$$

$$(\varepsilon + P(A))P(C_i \cup C_j) - P((C_i \cup C_j)A) < \varepsilon.$$

First, we want to list two facts about the restrictions Case 3 places on the values of atoms in our current partition. Let $\alpha_i = P(AC_i)$ and $\alpha_i^c = P(A^cC_i)$

FACT 1: *If* $B = \cup_{k\in Q} AC_k$ or $B = \cup_{k\in Q} A^cC_k$ for some $Q \subset \{1, \ldots, n\}$ such that $|Q| \geq 3$, then $P(B) \geq \frac{1}{2}|Q|$.

FACT 2: $\frac{1}{3} \leq P(A) < \frac{2}{3}$

Fact 1 holds because the minimum value for each α_i, $i \in Q$, is $\frac{1}{2}\varepsilon$. Otherwise, if $\delta = \frac{1}{2}\varepsilon - \alpha_i > 0$, then $\alpha_j \geq \frac{1}{2}\varepsilon + \delta$ for all $j \neq i, j \in Q$ to meet the requirement that $\alpha_i + \alpha_j \geq \varepsilon$. Then, $P(B) = \Sigma_{k\in Q}\, \alpha_k$ will be some positive multiple of δ greater than it might be if $\alpha_i = \frac{1}{2}$

Fact 2 holds by using Fact 1 on $Q = \{1, \ldots, n\}$ and remembering that each $\alpha_i < \varepsilon$ by assumption in Case 3.

Now we proceed to prove coarsening in Case 3 by splitting our work into two subcases, taking the easier case first. We assume that there is no coarsening of our $\{C_i\}$ partition which dilates A.

n is even: We assume that the inequalities (14) fail, hence "$\geq \varepsilon$" holds instead of "$< \varepsilon$" in (14). Add together the values on each respective side of these inequalities, whichever case holds, for pairs $C_1 \cup C_2, C_3 \cup C_4, \ldots C_{n-1} \cup C_n$. After some algebra we end up with (using that $C_1, \ldots, C_n$ is a partition):

$$\frac{n-2}{2}\varepsilon \leq \tag{15}$$

$$\sum_{S_P(A, C_i \cup C_{i+1}) > 1,\ i \text{ odd}} [(\alpha_i + \alpha_{i+1}) - (\textstyle\sum_{k=1}^{n} \alpha_k)(\alpha_i + \alpha_{i+1} + \alpha_i^c + \alpha_{i+1}^c)]$$

$$- \sum_{S_P(A, C_i \cup C_{i+1}) < 1,\ i \text{ odd}} [(a_i + \alpha_{i+1}) - (\textstyle\sum_{k=1}^{n}$$

$$\alpha_k)(\alpha_i + \alpha_{i+1} + \alpha_i^c + \alpha_{i+1}^c)].$$

Note that here, as in Case 2 of this theorem, we ignore the possibility that some $S_P(A, C_i \cup C_j) = 1$ because, if it happens, then the partition $\{C_i \cup C_j, (C_i \cup C_j)^c\}$ is a binary coarsening dilating A. Also note that the requirement that $S_P(A, C_i \cup C_j) > 1$ is identical to saying that the values in the square brackets in the first big sum of (15) is positive.

What we do now is show that substituting the highest allowable numbers into the right hand side of (15) is not enough to satisfy (15). Using Fact 1 and remembering that $2\varepsilon > \alpha_i + \alpha_{i+1} \geq \varepsilon$, we claim that the following equation is sufficient to derive a contradiction from:

$$\frac{n-2}{2}\varepsilon < p[2\varepsilon - y(3\varepsilon)] - q[\varepsilon - y(3\varepsilon)] \tag{16}$$

where $y = P(A)$, $p = |\{k | k \text{ odd and } S_P(A, C_k \cup C_{k+1}) > 1\}|$, and $q = |\{k | k \text{ odd and } S_P(A, C_k \cup C_{k+1}) < 1\}|$. Note that $p + q = \frac{n}{2}$. Also, $p, q \geq 1$ since otherwise, using Lemma 3.4, the entire probability space would not be independent of itself.

We see that $[2\varepsilon - y(3\varepsilon)]$ is maximal in (15)'s first square bracket by assuming that α_i and α_{i+1} are very close to ε and then $\alpha_i^c + \alpha_{i+1}^c$ is, at minimum, ε, giving us $y(3\varepsilon)$ is the second term. We take the order of terms to maximize that we do because Fact 2 tells us that y is between 1/3 and 2/3, making it more worthwhile to maximize the term not multiplied by y.

Similarly, we find that the second square bracket formula in (15) is minimized by the second square bracket in (16). Also, we see that we can replace the big sums in (15) by multiplication by p and q because the values in each pair $C_i \cup C_j$ might be independent of each other. Finally, we replace the "$\leq$" in (15) by a "$<$" in 16 since α_i and α_{i+1} can only be made close to ε.

We now want to maximize the right hand side of (16) by picking the best possible values for p and y. We do this by taking derivatives with respect to y and p and seeing how each can contribute to the maximum value of (16). It turns out that $\frac{n-2}{2} = p$ and $y = \frac{1}{3}$ are the optimal values. Substituting these values into (16) and doing some algebra gives $n - 2 > n - 2$.

n is odd: In the n even subcase, we did not use a joint restriction on the values of y and p which we will need in order to derive a contradiction in the present subcase. Otherwise the proof is similar, and we will skim over elements of the proof which do not change much.

Again, we assume that the inequality in (14) fails for all pairs $C_i \cup C_j$. As in the n even subcase, add together the resulting inequalities for $n - 1$ disjoint pairs, skipping some arbitrary element, say C_j—this produces an inequality similar to (15). Then repeat this process for each C_j and add all of those inequalities similar to (15) together. We end up with:

$$\frac{(n-2)(n-1)}{2}\varepsilon \leq \tag{17}$$

$$\sum_{l=1}^{n} \sum_{Sp(A,C_i \cup C_j)>l,\ i,j \neq l} [(\alpha_i + \alpha_j) - (\textstyle\sum_{k=1}^{n} \alpha_k)(\alpha_i + \alpha_j + \alpha_i^c + \alpha_j^c)]$$

$$-\sum_{l=1}^{n} \sum_{Sp(A,C_i \cup C_j)<l,\ i,j \neq l} [(\alpha_i + \alpha_j) - (\textstyle\sum_{k=1}^{n} \alpha_k)(\alpha_i + \alpha_j + \alpha_i^c + \alpha_j^c)]$$

where the indexes i and j make the inner sum cover each C_k once except for C_l.

Trying to maximize the values to make (17) true leads us as before to the inequality

$$\frac{(n-1)(n-2)}{2}\varepsilon < np[2\varepsilon - y(3\varepsilon)] - nq[\varepsilon - y(3\varepsilon)] \tag{18}$$

where $p + q = \frac{n-1}{2}$ with $p, q \geq 0$.

We now compute the aforementioned relation between the quantities p and y. We assume that we are given p and want to calculate the maximum value for y that p allows. Note that we do not need to go back and calculate (18) again because no variation in the value y makes it worthwhile to change the maximal values of the α_i's or α_i^c's which give us that inequality.

Using the values for the α_i's and α_i^c's in the calculation of (18) (as displayed in subcase n even), we find that

$$P(A) = y = \frac{P(A)}{P(A) + P(A^c)} = \frac{[2p\varepsilon + q\varepsilon + \frac{1}{2}\varepsilon]}{[2p\varepsilon + q\varepsilon + \frac{1}{2}\varepsilon] + [p\varepsilon + 2q\varepsilon + \varepsilon]}$$

Solving this as a function of p and n gives us

$$y = \frac{2p + n}{3n}$$

which means that the greater p is, the closer y must approach 2/3 for large n—something we did not need to use in the n even subcase.

Substituting this y (as well as $q = \frac{n-1}{2} - p$) into (18) allows us to compute a derivative in terms of p only. We find that the value $p = \frac{n-1}{2}$ maximizes (18) and by using this value in the inequality we derive:

$$\frac{(n-1)(n-2)}{2} < \frac{n-1}{2} \tag{18}$$

which is false for $n \geq 3$. RAA.

CASE 4: This case is a mixture of the conditions we used in Cases 2 and 3, as well as where half of the condition of Case 1 might hold. We conjecture that coarsening works in this hybrid case also, but the proof seems to be a very difficult one requiring inequalities to be derived for contradictions as in

Case 3, but without having the luxury of uniformity of conditions on the size of atoms and pairs of atoms relative to epsilon.

Appendix 2
Coarsening over Countable Partitions

The coarsening results we have derived (thus far) in this and a previous paper (Seidenfeld and Wasserman 1993) for the ε-contamination model, the ALUP model, the total variation model, and the fixed-marginals model have all critically used that the partition coarsened over has been discrete and finite. We now show that a coarsening result applies to the abovementioned models with countable, discrete partitions. We first consider the case where we have a countable, discrete partition of the ε-contamination model. We will derive the coarsening result for this model and then modify the proof as needed for the total variation, ALUP, and fixed-marginals models. In this subsection, the symbol N stands for the set of natural numbers.

We are given atoms a_{ij} for $i = 1, 2$ and $j \in N$ such that $B_i = a_{1i} \cup a_{2i}$ and $A = \cup_{j\in N} a_{1j}$. We also are given a probability distribution such that $P(a_{ij}) > 0$ for all i and j. We define $\alpha_A = P(\cup_{j\in N} a_{1j})$, $\alpha_A^c = P(\cup_{j\in N} a_{2j})$, and $\alpha_{ij} = P(a_{ij})$. For the model we are working in, the strict dilation conditions can be written as:

$$\textit{If } \delta(A, B) \geq 0 \textit{ then B dilates A iff } \delta(A, B) < \varepsilon\alpha_A(1 - \sum_{m\in N} (\alpha_{1a(m)}) + \alpha_{2a(m)})) \quad (19)$$

and

$$\textit{If } \delta(A, B) < 0 \textit{ then B dilates A iff } -\delta(A, B) < \varepsilon\alpha_{A^c}(1 - \sum_{m\in N} (\alpha_{1a(m)} + \alpha_{2a(m)})). \quad (20)$$

where we define $\delta(A, B) = (\Sigma_{m\in N}\alpha_{1a(m)})\alpha_{A^c} - (\Sigma_{m\in N}\alpha_{2a(m)})\alpha_A$ such that $B = \cup_{m\in N} B_{a(m)}$ for some index sequence a(m) (where the sequence has a finite or countable range).

THEOREM A2.1: *If the ε-contamination model with countable partition π' dilates A, then there exists a finite subpartition π of π' such that π dilates A also.*

COROLLARY A2.2: *If the ε-contamination model with countable partition π' dilates A, then there exists a binary subpartition π of π' such that π dilates A also.*

The corollary is proved using Theorem A2.1 and Theorem 3.1 of Seidenfeld and Wasserman 1993 where it is shown that every finite dilating partition has a binary coarsening which also dilates.

PROOF of A2.1: We will try to find an index sequence a(m) of natural numbers such that

1) there exists an n such that $a(n + i) = a(n) + i$ for all $i \in N$, and

2) $\{B = \cup_{m\in N} B_{a(m)}\} \cup \{B_i | i \in N - \{a(m) | m \in N\}\}$ is a subpartition dilating A.

We look at B_1 to find this index sequence. If $\delta(A, B_1) = 0$ then $\{B_1, B_1^c\}$ dilates A by (19) and (20) and, thus, we are done. If $\delta(A, B_1) \neq 0$, then without loss of generality assume that $\delta(A, B_1) > 0$.

We wish to choose a number q such that $a(1) = 1$ and $a(m + 1) = q + m - 1$ and such that $\delta(A, B) > 0$ in the new partition of 2). Note that 1) is already satisfied by the index sequence we selected. Thus, we need to see that the dilation condition for B in 2) holds:

$$\delta(A, B) < \varepsilon\alpha_A(1 - \sum_{m\in N}(\alpha_{1a(m)} + \alpha_{2a(m)})).$$

Expanding the left hand side and separating out one right hand side term we get:

$$(\alpha_{11} + \sum_{m\in N} \alpha_{1q+m-1})\alpha_{A^c} - (\alpha_{21} + \sum_{m\in N} \alpha_{2q+m-1})\alpha_A < \varepsilon\alpha_A(1 - \alpha_{11} - \alpha_{21}) - \varepsilon\alpha_A(\sum_{m\in N}(\alpha_{1q+m-1} + \alpha_{2q+m-1}))$$

This inequality is equivalent to:

$$\delta(A,B_1) + (\sum_{m\in N} \alpha_{1q+m-1})\alpha_{A^c} - (\sum_{m\in N} \alpha_{2q+m-1})\alpha_A < \varepsilon\alpha_A(1 - \alpha_{11} - \alpha_{21}) - \varepsilon\alpha_A(\sum_{m\in N}(\alpha_{1q+m-1} + \alpha_{2q+m-1})).$$

Since B_1 of the original partition dilates A, there exists, by (19), a positive ε_1 such that $\delta(A, B_1) = \varepsilon\alpha_A(1 - \alpha_{11} - \alpha_{21}) - \varepsilon_1$. We then transform the previous inequality into:

$$(\sum_{m\in N} \alpha_{1q+m-1})(\varepsilon\alpha_A + \alpha_{A^c}) + (\sum_{m\in N} \alpha_{2q+m-1})\alpha_A(\varepsilon - 1) < \varepsilon_1. \tag{21}$$

In (21) the second term on the left hand side is negative and $(\varepsilon\alpha_A + \alpha_{A^c}) < 1$, thus we only need to check that we can choose q such that $(\Sigma_{m\in N}\, \alpha_{1q+m-1}) < \varepsilon_1$. We can always find such a q because the sequence $\{\alpha_{1j}\}_{j\in N}$ converges.

Hence, we can collapse a tail sum of B_{q+i}'s (condition 1)) with B_1 in the original partition to obtain a finite subpartition which dilates A (condition 2)). □

Our remaining task is to say how to modify Theorem A2.1 (and thus Corollary A2.2) so that the countable to finite coarsening given in the proof of A2.1 works for ALUP, total variation, and fixed-margin models.

We consider the ALUP model countable partition coarsening. Using the notation from §2 of this paper, we recall that the dilation conditions replacing those of (19) and (20) are:

A is strictly dilated by $B = \cup_{m\in N} B_{a(m)}$ iff

$$\gamma_{A\cap B}\beta_{A^c} - \beta_{A^c\cap B}\gamma_A > 0 \textit{ and } \gamma_{A^c\cap B}\beta_A - \beta_{A\cap B}\gamma_{A^c} > 0.$$

We first need to prove a Lemma in order to justify A2.1 for the ALUP model.

LEMMA A2.3 *For an event E on the countable partition in the ALUP model,* $\beta_E \geq \beta_{E-a_{ij}}$ *(hence,* $\gamma_E \geq \gamma_{E-a_{ij}}$*) when* $a_{ij} \in E$ *and* E^c *is not empty.*

PROOF of A2.3: Assume the Lemma is false: $\beta_{E-a_{ij}} - \beta_E = \varepsilon > 0$ for some

E, some i,j. We select probability $p' \in [\beta_{E-a_{ij}}]$, and by Theorem 2.1 we can select a probability $p \in [\beta_E] \cap [\beta_{a_{ij}}]$ (the proof of Theorem 2.1 works even when E contains an infinite number of atoms). By our assumptions we have that $p'(a_{ij}) \geq p(a_{ij})$ and $p'(E - a_{ij}) = p(E - a_{ij}) + p(a_{ij}) + \varepsilon$. The latter equation shows that there exists k and l such that $p'(a_{kl}) = p(a_{kl}) + \varepsilon_1$ where $a_{kl} \in E - a_{ij}$ and $\varepsilon_1 > 0$. However, the former inequality implies that the additional mass that $E - a_{ij}$ has in p' but not in p must come from E^c, not a_{ij}, if we were to try to create p' from p. Hence, there must exist some r and s with $a_{rs} \in E^c$ such that $p(a_{rs}) = p'(a_{rs}) + \varepsilon_2$. But we can now create a new probability p'' in our given ALUP model from p' just by moving min $\{\varepsilon_1, \varepsilon_2\}$ mass from a_{kl} to a_{rs}. But then, $p''(E - a_{ij}) < p'(E - a_{ij})$, which contradicts our choice of p'. □

From the dilation conditions for the ALUP model we can see that in order to keep the left-hand side of the inequalities positive when coarsening B_1 with a tail sum of B_i's, we only need be able to choose a q such that, when $B = B_1 \cup B_q \cup B_{q+1} \cup \ldots$, $(\beta_{A\cap B} - \beta_{A\cap B_l})\,\gamma_{A^c} < \gamma_{A^c\cap B_l}$ and $(\beta_{A^c\cap B} - \beta_{A^c\cap B_l})\gamma_A < \gamma_{A\cap B_l}$. In other words, we do not expect any help from $\gamma_{A^c\cap B}$ or $\gamma_{A\cap B}$ other than that they are greater than $\gamma_{A^c\cap B_l}$ or $\gamma_{A\cap B_l}$, respectively, by Lemma A2.3. Thus, we only need to prove:

CLAIM: There exists, for any $\varepsilon > 0$, a q such that $\beta_{A\cap B} - \beta_{A\cap B_l} < \varepsilon$ (likewise for A^c).

PROOF: First, we note that $\lim_{q\in N} \beta_{a_{1q}} = 0$. Then, take any $p \in [\beta_{a_{11}}]$ and pick a q such that $p(A \cap (B_1 \cup B_q \cup B_{q+1} \cup \ldots)) < p(A \cap B_1) + \varepsilon$. By definition it must be that $\beta_{A\cap(B_l\cup B_q\cup B_{q+1}\cup\ldots)} < p(A \cap (B_1 \cup B_q \cup B_{q+1} \cup \ldots)) < p(A \cap B_1) + \varepsilon$.

Thus, we have shown how to prove the equivalent of Theorem A2.1 for the ALUP model with an infinite countable partition.

The modification of A2.1 for the total variations case is quite simple because it uses the same idea as Case 1 of the coarsening proof found in Appendix 1. Since the probability of the entire space is 1, we can find some $q \in N$ such that $P(B_q \cup B_{q+1} \cup B_{q+2} \cup \ldots) < \varepsilon$. This certainly implies that $P(AB_{q+i}), P(A^cB_{q+i}) < \varepsilon$ for all $i \in N$. By assuming that the infinite partition dilates A, we see by Lemma 3.2 case a) that $P(A), P(A^c) > \varepsilon$. However, this last inequality ensures that the event $B = \cup^{\infty}_{i=0} B_{q+i}$ dilates A also, again by Lemma 3.2 case a). Hence the finite partition $\{B_1, B_2, B_3, \ldots, B_{q-1}, B\}$ dilates A.

Finally, to prove the analogue of A2.1 for the fixed marginals model, we use any nontrivial, proper coarsening of the countable partition to a finite partition because the dilation conditions are easy to satisfy. The proof of coarsening here is the same as the proof for Theorem 4.3.

Appendix 3
Proof of Results

PROOF OF PROPOSITION 4:

Equation (5) implies that $(\underline{P}(A|B) - \underline{P}(A)) + (\overline{P}(A) - \overline{P}(A|B)) \geq 0$ with strict inequality for some B. This implies, after some algebra, that

$$\frac{\varepsilon(1 - \varepsilon)(1 - P(B))}{\varepsilon P(B) - P(B^c) - \varepsilon} \geq 0$$

which implies $P(B) \geq 1$. □

PROOF OF THEOREM 2.1:

We argue indirectly for $[\gamma_E] \cap [\gamma_{a_{ij}}] \neq \emptyset$. Assume that $max_{[\gamma_E]} P(a_{ij}) = \gamma_{a_{ij}} - \delta$ where $\delta > 0$. Let $q^* \in [\gamma_E]$ such that $q^*(a_{ij}) = \gamma_{a_{ij}} - \delta$. One of the following cases occurs:

(i) There exists some $\{a_{kl}\} \in E$ such that $q^*(a_{kl}) \neq \beta_{a_{kl}}$. Then we can shift some of the probability mass from a_{kl} to a_{ij}, without adding mass to event E, thereby contradicting the claim that $\gamma_{a_{ij}} - \delta$ is the maximum of a_{ij} over $[\gamma_E]$.

(ii) If for all $\{a_{kl}\} \in E$ we have that $q^*(a_{kl}) = \beta_{a_{kl}}$, then we have a contradiction with the requirement that $max_{\wp_{\beta\gamma}} P(a_{ij}) = \gamma_{a_{ij}}$ because getting any mass for a_{ij} would involve getting it from some $a_{kl} \in E^c$ thereby raising the probability for E past its maximum of γ_E.

The argument for β_E and $\beta_{a_{ij}}$ runs parallel to the one above. □

PROOF OF COROLLARY 2.2:

Let $E = X - \{a_{ij}\}$. Then for all $q \in [\gamma_E]$ we have $q(a_{ij}) = \beta_{a_{ij}}$ since $E^c = \{a_{ij}\}$. Then by Theorem 2.1 we see that there exists some $q^* \in [\gamma_E] \cap [\gamma_{a_{kl}}]$ which completes the corollary. □

PROOF OF COROLLARY 2.3:

(a) $\overline{P}(A) = 1 - \beta_{A^c}$ by definition, and $\overline{P}(A|b_j) = \dfrac{\gamma_{a_{1j}}}{\gamma_{a_{1j}} + \beta_{a_{2j}}}$ by Corollary 2.2. Then, we just compute the asserted result. Case (b) is similar. □

PROOF OF LEMMA 2.5:

We just need to set $\beta'_E = \min_{\wp} \Sigma_{a \in E} P(a)$, $\gamma'_E = \max_{\wp} \Sigma_{a \in E} P(a)$, and for all $a \in E^c \beta'_a = \beta_a$ and $\gamma'_a = \gamma_a$. It is easy to check that $\wp_E$ is an ALUP model if $\wp$ is one. □

PROOF OF LEMMA 2.6:

We will prove (1) by assuming that $\mathscr{G}$ contains no upper dilator and by deriving a contradiction ((2) is similar). Corollary 2.3 tells us that, if $\mathscr{G}$ contains no upper dilator,

$$\frac{\gamma_{A \cap E_i}}{\gamma_A} - \frac{\beta_{A^c \cap E_i}}{\beta_{A^c}} \leq 0 \text{ for all } E_i \in \mathscr{G}. \tag{22}$$

By Theorem 2.1 we may choose $P^{\gamma}_A \in [\gamma_A] \cap [\gamma_{a_{1k}}]$. When $b_k \in E_i$, choose $P^{\gamma}_{A \cap E_i} \in [\gamma_{A \cap E_i}] \cap [\gamma_{a_{1k}}]$. Otherwise, when $b_k \in E^c_i$, we let $P^{\gamma}_{A \cap E_i} \in [\gamma_{A \cap E_i}]$ be arbitrary.

By definition of $[\gamma_{A \cap E_i}]$ we have

$$\sum_{a_{1j} \in E_i \cap A} P^{\gamma}_{A \cap E_i}(a_{1j}) \geq \sum_{a_{1j} \in E_i \cap A} P^{\gamma}_A(a_{1j}). \tag{23}$$

Let $c + 1$ be the number of $E_i \in \mathscr{G}$ such that $b_k \in E_i$ ($c \geq 1$ by definition of a focused span). Then we compute:

$$\sum_{E_i \in \mathcal{G}} \gamma_{A\cap E_i} = (c + 1)\gamma_{a1k} + \sum_{j\neq k} P^{\gamma}_{A\cap E_i}(a_{1j}) \geq c\gamma_{a1k} + \sum_{j=1}^{n} P^{\gamma}_{A}(a_{1j}) = c\gamma_{a1k} + \gamma_A. \quad (24)$$

In (24), the first equation sign comes from the definition of the $P^{\gamma}_{A\cap E_i}$'s. The $\geq$ in (24) comes from (23). Finally, the last equality sign in (24) comes from the definition of γ_A.

By similar reasoning, we can derive an inequality for the lower ALUP bounds as:

$$\sum_{E_i \in \mathcal{G}} \beta_{A^c\cap E_i} \leq c\beta_{a2k} + \beta_{A^c}. \quad (25)$$

Using (24) and (25) we can derive that

$$\frac{\sum_{E_i\in\mathcal{G}} \gamma_{A\cap E_i}}{\gamma_A} - \frac{\sum_{E_i\in\mathcal{G}} \beta_{A^c\cap E_i}}{\beta_{A^c}} \geq \frac{c\gamma_{a1k} + \gamma_A}{\gamma_A} - \frac{c\beta_{a2k} + \beta_{A^c}}{\beta_{A^c}} = c\left(\frac{\gamma_{a1k}}{\gamma_A} - \frac{\beta_{a2k}}{\beta_{A^c}}\right) > 0. \quad (26)$$

The final inequality occurs because we assumed that b_k is an upper dilator for A. To obtain a contradiction we note that if we take the inequalities of (22) and sum them over all $E_i \in \mathcal{G}$, then, as no E_i is an upper dilator, the result is a formula which is identical to the first formula in (26), but where we end up with the sum being ≤ 0. □

PROOF OF LEMMA 2.7:

We will prove (1) by assuming the opposite and deriving a contradiction. Thus, we have, by Corollary 2.3

$$\frac{\gamma_{A\cap E^j}}{\gamma_A} - \frac{\beta_{A^c\cap E^j}}{\gamma_{A^c}} \leq 0 \text{ for } j = 1, \ldots, n. \quad (27)$$

If we sum (27) over j and add and subtract $n - 1$ we derive

$$\frac{\sum_{j=1}^{n} \gamma_{A\cap E^j} - (n-1)\gamma_A}{\gamma_A} + \frac{(n-1)\beta_{A^c} - \sum_{j=1}^{n} \beta_{A^c\cap E^j}}{\beta_{A^c}} \leq 0. \quad (28)$$

However, since $(n - 1)\gamma_A = \sum_{j=1}^{n} \sum_{i=1, i\neq j}^{n} P^{\gamma}_{A}(a_{1i})$ (and similarly for $(n - 1)\beta_{A^c}$), we can rewrite (28) as

$$\frac{\sum_{j=1}^{n} (\gamma_{A\cap E^j} - \sum_{i=1,\, i\neq j}^{n} P^{\gamma}_{A}(a_{1i}))}{\gamma_A} + \frac{\sum_{j=1}^{n} (-\beta_{A^c\cap E^j} + \sum_{i=1, i\neq j}^{n} P^{\beta}_{A}(a_{2i}))}{\beta_{A^c}} \leq 0. \quad (29)$$

On the other hand we notice that for any probability distribution P we have $\sum_{i=1, i\neq j}^{n} P(a_{1i}) \leq \gamma_{A\cap E^j}$ and $\sum_{i=1, i\neq j}^{n} P(a_{2i}) \geq \beta_{A^c\cap E^j}$. In particular, let $P \in [\gamma_A]$ in the first inequality of the previous sentence and then $P \in [\beta_{A^c}]$ in the second inequality above. Then the formulas in the scope of the "j" summands of (29) are always non-negative. But the only way that (29) is possible, then, is if

$$\sum_{i=1, i\neq j}^{n} P_A^{\gamma}(a_{1i}) = \gamma_{A\cap E^j}, \sum_{i=1, i\neq j}^{n} P_A^{\beta}(a_{2i}) = \beta_{A^c\cap E^j} \text{ for all } j = 1, \ldots, n. \quad (30)$$

We can think of (30) as n linearly independent equations with n unknowns, which are the $P_A^{\gamma}(a_{1i})$ (similarly for β). Hence, $[\gamma_A]$ forces unique values for each $(P_A^{\gamma}(a_{1i}))$. But by Theorem 2.2, for all i there exists a $P \in [\gamma_A]$ such that $P(a_{1i}) = \gamma_{a_{1i}}$. Hence, $\gamma_A = \Sigma_{i=1}^{n} \gamma_{a_{1i}}$. We can then see that

$$\gamma_{A\cap E^j} = \sum_{i=1, j\neq i}^{n} \gamma_{a_{1i}} \text{ and } \beta_{A^c\cap E^j} = \sum_{i=1, j\neq i}^{n} \beta_{a_{2i}} \quad (31)$$

because $A = (A \cap E^j) \cup (A \cap b_j)$. Finally, we obtain a contradiction by using the assumption of Theorem 2.4 and Corollary 2.3 to obtain:

$$\frac{\gamma_{a_{1i}}}{\gamma_A} - \frac{\beta_{a_{2i}}}{\beta_A} > 0$$

Summing over $i = 2, \ldots, n$ and using (31) we get:

$$\frac{\gamma_{A\cap E^1}}{\gamma_A} - \frac{\beta_{A^c\cap E^1}}{\beta_{A^c}} > 0.$$

This contradicts (27). □

PROOF OF LEMMA 2.8:

We will prove (1) by assuming that $\mathcal{F}$ is a tiling of X without an upper dilator. By Lemma 2.7 we know that some E^k is a lower dilator. This E^k is not an upper dilator also because, if it were, $\{E^k, b_k\}$ would be a coarsened subpartition of the probability space which dilates A (i.e., Theorem 2.4 would be true). Pick out $E_{jk} \in \mathcal{F}$ (it is unique because F is a tiling). Then, $S = \{E^k, E_{jk}\}$ spans X with focus b_j, and S contains no upper dilator. However, this directly contradicts Lemma 2.6. □

PROOF OF LEMMA 2.9:

Assume we have a covering of $\mathcal{F}$ using labels U, L, and N obeying 2.6 and 2.8's restrictions. 2.6 and 2.8 say that there cannot be a focused span or tiling in $\mathcal{F}$ consisting of all not-L labels or all not-U labels. Thus, if we arbitrarily change the N labels to U or L labels, we will have fewer restrictions of 2.6 and 2.8 to obey. □

PROOF OF THEOREM 2.4:

We argue by assuming, in order to derive a contradiction, that there exists a labeling of $\mathcal{F}$ containing only U's and L's. This is sufficient given Lemma 2.9. Define $r_L = \min_i$ (number of L labels in $\mathcal{F}_i$) and $r_U = \min_i$ (number of U labels in $\mathcal{F}_i$). Note that Lemma 2.8 applies to each $\mathcal{F}_i$ so $r_L, r_U \geq 1$. Without loss of generality assume that $r_L \leq r_U$. Pick a $\mathcal{F}_i'$ such that (the number of L's in $\mathcal{F}_{i'}$) $= r_L$. Then, rearrange the indices of the b_j so that the current $\mathcal{F}_{i'}$ becomes $\mathcal{F}_1$ and such that

$$\mathcal{G}_1 = \{E_{12}, E_{13}, \ldots, E_{1m+1}, E_{1m+2}, \ldots, E_{1n}\}$$

where all of the E_{1j}, $j \leq m + 1$ are labeled U, and E_{1j}, $j > m + 1$ are labeled L.

For a contradiction we will pick out a spanning set S_m of E_{ij} pairs which either focuses on b_1 or is a tiling and such that all the E_{ij} are labeled U. We define $E_{i_j j} \in S_m$ for decreasing j as follows:

1) j = n: From $\mathcal{G}_n$ choose $E_{i_n n} \in S_m$ such that $i_n = \max_i(E_{in}$ is labeled U).

2) $n > j \geq m + 2$:

a) If there exists $l > j$ such that $i_l = j$ then choose $E_{i_j j} \in S_m$ and $i_j = l$.

b) Otherwise choose $E_{i_j j} \in S_m$ where $i_j < j$ and $i_j = \max_i(E_{ij}$ is labeled U and $i \neq i_l$ for all $j + 1 \leq l \leq n$).

By the case we are in, the construction is possible since for $i > 1$ there are at least r_U $E_{ij} \in \mathcal{G}_i$ labeled U, and $r_U \geq r_L = n - m - 1 =$ the number of $E_{i_l l}$'s we need to put into S_m by our construction.

Fact: No $i_L = 1$ in the construction since $l \geq m + 2$, and we assumed that we rearranged the indices to get E_{1l} labeled L.

Define $J = \{1 < j \leq m + 1 | j \neq i_k \text{ for all } m + 2 \leq k \leq n\}$.

CLAIM: $J \neq \emptyset$

PROOF: First, from the case we are in we have $n - m - 1 = r_L \leq (n - 1)/2$, and since $r_L =$ the number of L labels in $\mathcal{G}_1$. Thus, the number of U labels in $\mathcal{G}_1 = m \geq (n - 1)/2$. Putting these together we get

$$m \geq (n - 1)/2 \geq n - m - 1. \tag{32}$$

We saw that n-m-1 is the number of indices i_l we defined in our construction of S_m, and $|J| \geq m - (n - m - 1)$. If n is even we see that $(n - 1)/2$ is fractional, thus (32) becomes $m > (n - 1)/2 > n - m - 1$, and we end up with $|J| \geq 1$. In the case when n is odd, we see as before that $J \neq \emptyset$ unless, possibly, $m = (n - 1)/2 = n - m - 1 = r_L$, and section 2) a) in our construction of S_m is never used. However, by our Fact, we deduce that in this case there are only $(n - 1)/2 - 1 = m - 1$ columns in our table F to be matched up with the $n - m - 1 = m$ indices i_k that are defined in the construction. Thus, 2) a) must be used once in the construction. *end of claim*

We now add $\cup_{j \in J}\{E_{1j}\}$ to S_m. If $|J| = 1$, then S_m is a tiling containing only E_{ij}'s labeled U. However, if $|J| \geq 2$, S_m is a span focused at b_1 containing only E_{ij}'s labeled U. □

PROOF OF LEMMA 2.11:

From the proof of Corollary 2.3 we have: $\Delta(A, \{b_j\}) = \left(\dfrac{\gamma_{a_{1j}}}{\gamma_{a_{1j}} + \beta_{a_{2j}}} - \gamma_A\right) - \left(\beta_A - \dfrac{\beta_{a_{1j}}}{\beta_{a_{1j}} + \gamma_{a_{2j}}}\right)$. After this the Lemma follows by straightforward computation using that the probabilities P_{1j} and P_{2j} can exist by 2.1 and 2.2. □

PROOF OF COROLLARY 2.12:

$Cov(\chi_A, \chi_{b_j}) = E_P\{(\chi_A - P(A))(\chi_{b_j} - P(b_j))\} = E_P\{\chi_{a_{1j}} - P(A)P(b_j)\} = P(a_{1j}) - (\Sigma_{i=1}^n P(a_{1i}))(P(a_{1j}) + P(a_{2j})) = \delta_P(A, b_j)$ after an easy calculation. □

PROOF OF LEMMA 3.2:

1. Assume that $P(AB)$, $P(A^cB) < \varepsilon$. Then, from (6) we see that B strictly dilating A is equivalent to a. and b.:

a. $\underline{P}(A) > \underline{P}(A|B)$ iff

$$\max\{P(A) - \varepsilon, 0\} > \frac{\max\{0, P(AB) - \varepsilon\}}{\max\{0, P(AB) - \varepsilon\} + \min\{1, P(A^cB) + \varepsilon\}}.$$

iff $P(A) > \varepsilon$.

b. $\overline{P}(A) < \overline{P}(A|B)$ iff

$$\min\{P(A) + \varepsilon, 1\} < \frac{\min\{P(AB) + \varepsilon, 1\}}{\min\{P(AB) + \varepsilon, 1\} + \max\{0, P(A^cB) - \varepsilon\}}$$

iff $1 - P(A) = P(A^c) < \varepsilon$. Thus, part 1. of the Lemma holds.

2. When $P(AB) < \varepsilon \leq P(A^cB)$ strict dilation is equivalent to c. and d.:

c. $\underline{P}(A) > \underline{P}(A|B)$ iff $P(A) > \varepsilon$ as in a.

d. $\overline{P}(A) < \overline{P}(A|B)$ iff

$$\min\{P(A) + \varepsilon, 1\} < \frac{\min\{P(AB) + \varepsilon, 1\}}{\min\{P(AB) + \varepsilon, 1\} + \max\{0, P(A^cB) - \varepsilon\}}$$

iff

$$P(A) + \varepsilon < \frac{P(AB) + \varepsilon}{P(AB) + \varepsilon + P(A^cB) - \varepsilon}$$

because the assumption that $P(A^c) \geq P(A^cB) \geq \varepsilon$ makes $P(A) + \varepsilon > 1$ and $P(AB) + \varepsilon > 1$ impossible when evaluating the max and min in the equation. Equivalently, we have $\varepsilon > (P(A) - P(AB)/P(B))/((1/P(B)) - 1)$ iff $\varepsilon > -d_P(A, B)/P(B^c)$ iff $\varepsilon > P(A^c)(1 - S_P(A^c, B^c))$. Hence part 2, of the Lemma holds.

3. When $P(A^cB) < \varepsilon \leq P(AB)$ strict dilation is equivalent to e. and f.:

e. $\underline{P}(A) > \underline{P}(A|B)$ iff $P(A)(1 - S_P(A, B^c)) < \varepsilon$ by a proof analogous to that given in d.

f. $\overline{P}(A) < \overline{P}(A|B)$ iff $\varepsilon < P(A^c)$ by the same proof given in b.

4. When $\varepsilon \leq P(AB), P(A^cB)$, the Lemma holds by the proofs in parts d. and e. □

Proof of Theorem 3.5:

We assume that the theorem is false and derive various contradictions. Thus, we assume that we are given a partition $\pi_n = \{C_1, C_2, \ldots, C_n\}$ which dilates the event A such that no coarsening $\{B_1, B_2, \ldots, B_m\}$ of π_n dilates A.

We see from Case 4 of Lemma 3.2 that independence of any of the C_i with respect to A is sufficient to cause the subpartition $\{C_i, C_i^c\}$ to strictly dilate A. So assume that none of the C_i are independent of A (under P) and let $\pi_n = C^+ \cup C^-$ where $C^+ = \{C_i \in \pi_n | S_P(A, C_i) > 1\}$ and $C^- = \{C_i \in \pi_n | S_P(A, C_i) < 1\}$. W.l.o.g. assume that $|C^-| \geq |C^+|$ and define $k = |C^+|$. Also, assume that $C^+ = \{C_1, \ldots, C_k\}$ and $C^- = \{C_{k+1}, \ldots, C_n\}$ and define $E_{i,j} = \cup\{C_k \in (\pi_n - \{C_i, C_j\})\}$ such that $C_i \in C^+$ and $C_j \in C_-$.

SUBCASE A: There exists $i, j \in \{1, \ldots, n\}$ s.t. $S_P(A, E_{i,j}) < 1$. From Lemmas 3.3 and 3.4 we see that $S_P(A, C_i \cup C_j) > 1$ because $E_{i,j}^c = C_i \cup C_j$. Using the assumption that the subpartition $\{C_j, C_j^c\}$ cannot strictly dilate A, Lemma 3.2 and the fact that all atoms have mass at least ε imply that

$$\varepsilon \leq P(A)(1 - S_P(A, C_j)) \textit{ because } C_j \in C^-. \quad (1)$$

Let $\pi_{ij} = \{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_{j-1}, C_{j+1}, \ldots, C_n, C_i \cup C_j\}$. Since this subpartition cannot strictly dilate A, we can use Lemma 3.2 to see that

$$\varepsilon \leq P(A)(1 - S_P(A, E_{i,j})) \textit{ using subcase A's assumption.} \quad (2)$$

We can use Lemma 3.4 on (1) and (2) to come up with:

$$\varepsilon \leq P(A)(1 - S_P(A, E_{i,j} \cup C_j)) = P(A)(1 - S_P(A, C_i^c)). \quad (3)$$

However, (3) contradicts the initial assumption we made that π_n, and in particular C_i, dilates A using Lemma 3.2 and that $C_i \in C^+$.

SUBCASE B: We have $S_P(A, E_{i,j}) > 1$ for all possible i, j. Note that the other possibility, that $S_P(A, E_{i,j}) = 1$, cannot happen because if it did the subpartition $\{E_{i,j}^c, E_{i,j}\}$ would strictly dilate A. The Lemmas 3.3 and 3.4 say that $S_P(A, E_{i,j}^c) < 1$ where $E_{i,j}^c = C_i \cup C_j$ for all $i \in \{1, \ldots, k\}$ and $j \in \{k + 1, \ldots, n\}$. Define $F_i = C_i \cup C_{i+k}$ *for* $i = 1, \ldots, k$. This implies that $S_P(A, F_i) < 1$ for all $i = 1, \ldots, k$. If we define $F = \cup_{i=1}^k F_i$, then $S_P(A, F) < 1$ by Lemma 3.4. Finally, since any element of π_n in the complement of F (if such elements exist) are elements of C^-, we can use Lemmas 3.3 and 3.4 to see that $S_P(A, F^c) < 1$, hence, $S_P(A, F) > 1$. But the final inequality contradicts one a few lines up. □

PROOF OF COROLLARY 3.7:

The corollary for cases a. and d. is by simple calculation, as is $\Delta(A, \{B_i\}) \geq 0$ oftentimes being equivalent to B_i dilating A. □

PROOF OF LEMMA 4.1:

To compute $\underline{P}(A|B_i)$ we note that the probability distribution $P_{min} \in \wp$ such that $P_{min}(AB_i) = max\{0, r + c_i - 1\}$ and $P_{min}(AB_j) = c_j$, $(i \neq j)$ provides the least mass possible on AB_i. Similarly, $P_{max} \in \wp$ such that $P_{max}(AB_i) = min\{r, c_i\}$ provides the greatest mass on AB_i possible so that $\overline{P}(A|B_i)$ is as in (7). □

PROOF OF THEOREM 4.2:

By the model and Lemma 4.1 we see that $\underline{P}(A > \underline{P}(A|B_i)$ if and only if (two cases): $c_i r > max\{0, r + c_i - 1\}$ iff

case 1: $rc_i > 0$ if $0 > r + c_i - 1$.

case 2: $r(c_i - 1) > c_i - 1$ if $r + c_i - 1 \geq 0$.

The condition in either case holds iff r and c_i are anything except 0 and 1.

Also, $\overline{P}(A) < \overline{P}(A|B_i)$ iff $rc_i < min\{r, c_i\}$ which is the case iff r or c_i are not 0 or 1 because $0 \leq r, c_i \leq 1$. □

PROOF OF THEOREM 4.3:

Because each B_i strictly dilates A, Theorem 4.2 says that any coarsening including B_i will have a marginal value of $c \geq c_i > 0$. Also, the marginal value of A, r, stays fixed under the coarsening of a partition; hence we can apply Theorem 4.2 again to get our result if we remember that $c \neq 1$ in a nontrivial subpartition of B. □

PROOF OF PROPOSITION 4.4: (Sketch)

As in Lemma 4.1, to calculate $\underline{P}(A|B)$ we need to find a distribution F that minimizes the mass on AB:

$$\underline{P}(A|B) = \frac{\min_F\{F(b,d) + F(a,c) - F(b,c) - F(a,b)\}}{F_Y(d) - F_Y(c)}$$

But we note that since for all x, y, $\min\{F_X(x), F_Y(y)\} \geq F(x,y) \geq \max\{0, F_X(x) + F_Y(y) - 1\}$ (Dall'Aglio 1972), the calculation is nearly identical to that in 4.1 and 4.2. It appears that 4.3 holds for any open sets A and B.

PROOF OF LEMMA 5.1.2:

Let $z = \mu(A)$: Without loss of generality assume that $A = [0, z]$. Let $t_B = \mu(A \cap B)$ and $u_B = \mu(A^c \cap B)$. Balance implies that $t_B/z = u_B/(1 - z)$ which implies that $u_B/t_B = (1 - u_B - z)/(z - t_B)$. Choose $f \in \langle A \rangle_0$ and let $F(\omega) = \int_0^\omega f_*(t)\mu(dt)$. Now,

$$\frac{1 - F(1 - u_B)}{F(t_B)} = \frac{\int_{1-u_B}^1 f_*}{\int_0^{t_B} f_*} \geq \frac{u_B f_*(1 - u_B)}{t_B f_*(t_B)} = \frac{(1 - u_B - z)f_*(1 - u_B)}{(z - t_B)f_*(t_B)}$$
$$\geq \frac{\int_z^{1-u_B} f_*}{\int_{t_B}^z f_*} = \frac{F(1 - u_B) - F(z)}{F(z) - F(t_B)}. \quad (33)$$

By Lemma 5.1.1, with $a_1 = t_B$, $a_2 = z$, $a_3 = 1 - u_B$, we have $F(t_B)/(F(t_B) + 1 - F(1 - u_B)) \leq F(z)$. So,

$$\underline{P}(A|B) \leq \inf\{P(A|B);\, dP/d\mu \in \Lambda(f)\} \leq F(t_B)/(F(t_B) + 1 - F(1 - u_B)) \leq F(z) = \underline{P}(A).$$

By a similar argument, $\overline{P}(A|B) \geq \overline{P}(A)$. □

PROOF OF LEMMA 5.1.3:

Choose $t \in (0, z)$ and let $u = (1 - z)t/z$. Thus, $u \in (0, 1 - z)$ and $t/z = u/(1 - z)$. Let $B = [0, t] \cup [1 - u, 1]$. Then $\mathcal{B} = \{B, B^c\}$ is balanced. Now apply the previous Lemma. □

PROOF OF LEMMA 5.1.4:

Suppose that $\phi_0(A)\phi^0(A) > 1$. Let $\mathcal{B} = \{B, B^c\}$ be a weak, balanced, binary dilator. It can be shown that at least one inequality in (33) is strict. □

PROOF OF LEMMA 5.1.5:

The "if" direction follows from 5.1.4. Conversely, suppose that $\phi_0(A)\phi^0(A) = 1$. Without loss of generality assume that m is convex and closed with respect to the total variation topology (otherwise replace it with its convex closure and none of the upper and lower probabilities change). As shown in Wasserman and Kadane 1992, since $\overline{P}$ is 2-alternating and m is I-closed, there exists a density f such that m is the closed, convex hull of $\{g;\, g \prec f\}$. Let $z = \mu(A)$, $t_i = \mu(A \cap B_i)$ and $u_i = \mu(A^c \cap B_i)$. From the form of m we have $\underline{P}(A) = \int_0^z f_*$ and $\underline{P}(A|B_i) = \int_0^{t_i} f_* / (\int_0^{t_i} f_* + \int_{1-u_i}^1 f_*)$. Dilation of the lower probability implies, by Lemma 5.1.1, that

$$\frac{\int_{1-u_i}^1 f_*}{\int_0^{t_i} f_*} \geq \frac{\int_z^{1-u_i} f_*}{\int_{t_i}^z f_*}.$$

If f_* is constant almost everywhere on A and A^c then the above inequality implies that $u_i/t_i \geq (1 - u_i - z)/(z_i - t)$ i.e. $zu_i \geq t_i(1 - z)$. Summing over i and noting that $\Sigma t_i = z$, $\Sigma u_i = 1 - z$ and that at least one of the inequalities is strict, we get $z(1 - z) > z(1 - z)$, a contradiction. So f_* is not constant over at least one of A or A^c implying that $\phi_0(A)\phi^0(A) > 1$. □

PROOF OF LEMMA 5.1.6:

Constriction implies $\underline{P}(A) \leq \underline{P}(A|B_i)$ for all i, strict for some i. This implies

$$\int_{t_i}^{z} f_* \int_{1-u_i}^{1} f_* \leq \int_{z}^{1-u_i} f_* \int_{0}^{t_i} f_* \tag{34}$$

where $t_i = \mu(A \cap B_i)$, $u_i = \mu(A^c \cap B_i)$, $z = \mu(A)$, $\Sigma t_i = z$ and $\Sigma u_i = 1 - z$. Now, since f_* is non-decreasing, $(z - t_i)u_i f_*(t_i)f_*(1 - u_i)$ is less than or equal to the left hand side of (34). Also, $(1 - u_i - z)t_i f_*(t_i)f_*(1 - u_i)$ is greater than or equal to the right hand side of (34). Thus, $u_i(z - t_i) \leq t_i(1 - u_i - z)$ for all i and summing over i and noting that the inequality is strict for some i we get $(1 - z)z < (1 - z)z$, a contradiction. □

PROOF OF LEMMA 5.2.1:

Let $z = \mu(A)$ and assume without loss of generality that $A = [0, z]$. It is easy to see that for any P and B, since B belongs to a nontrivial partition, $P(A|B) \geq f_\square$ where $f = dP/d\mu$. Thus, $\underline{P}(A|B) \geq m_\square$ and $\max_{B\in\mathcal{B}} \underline{P}(A|B) \geq m_\square$. Hence, $\inf_{\mathcal{B}\in\Gamma} \max_{B\in\mathcal{B}} \underline{P}(A|B) \geq m_\square$. Next we show that we can get arbitrarily close to this lower bound.

Let $\varepsilon > 0$ and choose $g \in m$ such that $m_\square \leq g_\square < m_\square + \varepsilon$.

Define $C_i = [(i - 1)z/n, iz/n)$ for $i = 1, \ldots, n$ and $D_i = [1 - i(1 - z)/n, 1 - (i - 1)(1 - z)/n)$ for i; eq $2, \ldots, n$ and $D_1 = [1 - (1 - z)/n, 1]$. Set $B_i = C_i \cup D_i$ and let $\mathcal{B}_n = \{B_1, \ldots, B_n\}$. Then $\min_i \underline{P}(A|B_i) = \underline{P}(A|B_1)$. Hence,

$$\max_{B\in B_n} \underline{P}(A|B) = \underline{P}(A|B_1) = \inf_f \frac{\int_0^{1/n} f_*}{\int_0^{1/n} f_* + \int_{(n-1)/n}^1 f_*} \leq \frac{\int_0^{1/n} g}{\int_0^{1/n} g + \int_{(n-1)/n}^1 g}.$$

The last expression tends to $g_\square < m_\square + \varepsilon$ as $n \to \infty$. A similar proof can be used for the other bound. □

PROOF OF LEMMA 6.1:

Since $\mu \in M$,

$$\begin{aligned}\underline{P}(A_n|X_n = x_n) &\leq \mu(A_n|X_n = x_n)\\ &= \frac{\Phi(\sqrt{n}(a_n - x_n)) - \Phi(\sqrt{n}(-a_n - x_n))}{\Phi(\sqrt{n}(1 - x_n)) - \Phi(\sqrt{n}(-x_n))} \to 0 \ a.s.\end{aligned}$$

where Φ is the cumulative distribution function for the standard normal.

PROOF OF LEMMA 6.3:

From a Theorem in Lavine, Wasserman and Wolpert 1993, it follows that $\underline{P}(A_n|X_n = x_n)$ is the solution to the equation $s(\lambda) = 0$, $0 \leq \lambda \leq 1$, where $s(\lambda) = \inf_{P\in M} \int h_\lambda \, dP$, $h_\lambda(\omega) = L(\omega)(I_{A_n}(\omega) - \lambda)$ and L is the likelihood function. Here, I_A represents the indicator function for the set A. So it suffices to identify, for every λ, the member $P \in M$ that minimizes $\int h_\lambda \, dP$. And for this, it clearly suffices to find the member $r \in S$ that minimizes $\int h_\lambda r_n d\mu$.

Define $r_n(\omega) = q_\alpha(\mu(h_\lambda(t) > h_\lambda(\omega)))$ where $q_\alpha(\theta) = (1 - \alpha)(1 - \theta)^{-\alpha}$ is the increasing rearrangement of p_α. By construction, $r_n \sim p_\alpha$. There exists b such that $g = h_\lambda + b$ is nonnegative. Let $z = \sup g < \infty$, let $Q_\alpha(d\omega) = q_\alpha(\omega)\mu(d\omega)$

and let $R_n(d\omega) = r_n(\omega)\mu(d\omega)$. Since q_α is the increasing rearrangement, we have that, for any P such that $dP/d\mu \in \mathcal{S}$, $\int h_\lambda\, dP = \int g\, dP - b = \int_0^z P(g > t)dt - b \geq \int_0^z Q_\alpha(g > t)dt - b = \int_0^z R_n(g > t)dt - b = \int g\, dR_n - b = \int h_\lambda\, dR_n$ This confirms that R_n is the appropriate minimizer. Finally, note that by the definition of r_n, the stated conditions on r_n follow for all λ. □

PROOF OF LEMMA 6.4:

After some algebra applied to 6.3, we have that

$$\underline{P}(A_n|X_n = x_n) = \frac{(1 - \varepsilon)\mu(A_n|X_n = x_n)m(x_n) + \varepsilon \int_{A_n} L_n r_n}{(1 - \varepsilon)m(x_n) + \varepsilon \int L_n r_n}$$

where $m(x_n) = \int L_n(\theta)\mu(d\theta) \to 1$. From (11), $\mu(A_n|X_n = x_n)$ stays bounded away from 0. Similarly, from (11), $\int_{A_n} L_n r_n \leq \sup_{A_n} r_n = 1 - \alpha < \infty$. □

PROOF OF LEMMA 6.5:

It suffices to consider $I_n = \int_{x_n+a_n}^1 L_n(\theta)r_n(\theta)d\theta$ since the integral over A_n is $O(1)$ and the integral to the left of $x_n - a_n$ behaves the same as I_n. Changing variables we have

$$\begin{aligned} I_n &= \sqrt{n}\int_{x_n+a_n}^1 \exp\{-n(\theta - x_n)^2/2\}(\theta - x_n - a_n)^{-\alpha}d\theta \\ &= \sqrt{n}\int_0^{1-x_n-a_n} \exp\{-n(u + a_n)^2/2\}u^{-\alpha}du \\ &= \sqrt{n}\exp\{-na_n^2/2\}\int_0^{1-x_n-a_n} \exp\{-nu^2/2\}\exp\{-nua_n\}u^{-\alpha}du. \end{aligned}$$

By another change of variables ($t = nu^2/2$) I_n is proportional to

$$J_n = exp\{-na_n^2/2\}n^{\alpha/2}\int_0^\infty I_{[0,n(1-x_n-a_n)^2/2]}(t)\exp\{-a_n\sqrt{2nt}\}e^{-t}t^{(1-\alpha)/2-1}dt.$$

Now suppose that (13) holds. So the inequality is strict for some $k > 0$. It follows that $\lim\,[\alpha/2 - na_n^2/(2\log n) - k\sqrt{n}a_n/(2\log n)] > 0$ so that

$$\lim n^{\alpha/2}e^{-na_n^2/2}e^{-\sqrt{2bn}a_n} = \infty$$

where $b = k^2/8 > 0$. Since $X_n \to \theta$ almost surely and $a_n = o(1)$, we conclude that $\lim n(1 - X_n - a_n)^2/2 \geq b$ a.s. so that

$$J_n \geq \lim n^{\alpha/2}e^{-na_n^2/2}e^{-\sqrt{2bn}a_n}z \text{ a.s.}$$

where $z = \int_0^b e^{-t}t^{(1-\alpha)/2-1}dt > 0$. It follows that $\lim J_n = \infty$ almost surely. □

PROOF OF THEOREM 6.6:

If $d < \alpha$ then (13) holds and the conclusion follows from Lemma 6.5. If $d \geq \alpha$ then $\lim n_n^{\alpha/2}e^{-na^2}/2 > \infty$. Since, $J_n \leq C_n = exp\{-na_n^2/2\}n^{\alpha/2}\Gamma((1 - \alpha)/2)$, the conclusion follows. □

REFERENCES

Berger, J. (1984), "The robust Bayesian viewpoint" (with discussion), in J. Kadane (ed.), *Robustness in Bayesian Statistics.* Amsterdam: North-Holland, pp. 63–124.

———. (1985), *Statistical Decision Theory* (2nd Edition). New York: Springer-Verlag.

———. (1990), "Robust Bayesian analysis: sensitivity to the prior", *J. Statist. Plann. Inference* 25: 303–328.

Berger, J. and M. Berliner (1986), "Robust Bayes and empirical Bayes analysis with epsilon-contaminated priors", *Ann Statist.* 14: 461–486.

Blackwell, D. and L. Dubins (1962), "Merging of opinions with increasing information", *Ann. Statist.* 33: 882–887.

Dall'Aglio, G. (1972), "Frechet Classes and Compatibility of Distribution Functions", *Symposia Mathematica* 9: 131–150. Providence, RI: Amer. Math. Soc.

DeRobertis, L. and J. A. Hartigan (1981), "Bayesian inference using intervals of measures", *Ann. Statist.* 9: 235–244.

Fine, T. L. (1988), "Lower probability models for uncertainty and nondeterministic processes", *J. Stat. Planning and Inference* 20: 389–411.

Good, I. J. (1952), "Rational decisions", *J. Roy. Statist. Soc. B.* 14: 107–114.

Huber, P. J. (1981), *Robust Statistics.* New York: Wiley.

Huber, P. J. and V. Strassen (1973), "Minimax tests and the Neyman-Pearson lemma for capacities", *Ann. Stat.* 1: 251–263.

Herron, T., T. Seidenfeld, and L. Wasserman (1994), "The extent of dilation of sets of probabilities and the asymptotics of robust Bayesian inference", in D. Hull, M. Forbes, and R. M. Burian (eds.), *PSA-1994,* vol 2. East Lansing: PSA, pp. 250–259.

Jeffreys, H. (1961), *Theory of Probability.* Oxford: Clarendon Press.

Kyburg, H. (1961), *Probability and the Logic of Rational Belief.* Middleton, CT: Wesleyan University Press.

———. (1974), *The Logical Foundations of Statistical Inference.* Dordrecht: Reidel.

Lavine, M. (1991), "Sensitivity in Bayesian statistics: the prior and the likelihood," *J. Amer. Statist. Assoc.* 86: 396–399.

Lavine, M., L. Wasserman, and R. Wolpert (1993), "Linearization of Bayesian robustness problems", *J. Statist. Plann. and Inference* 37: 307–316.

Levi, I. (1974), "On indeterminate probabilities", *J. Phil.* 71: 391–418.

———. (1980), *The enterprise of knowledge.* Cambridge, MA: MIT Press.

———. (1982), "Conflict and social agency", *J. Phil.* 79: 231–247.

Pericchi, L. R. and P. Walley (1991), "Robust Bayesian credible intervals and prior ignorance", *Internat. Statist. Review* 59: 1–24.

Rachev, S. T. (1985), "The Monge-Kantorovich mass transference problem and its stochastic applications", *Theory of Probability and Its Applications* 29: 647–671.

Savage, L. J. (1972), *The Foundations of Statistics,* (2nd Edition). New York: Dover.

Schervish, M. and T. Seidenfeld (1990), "An approach to consensus and certainty with increasing shared evidence", *J. Stat. Planning and Inference* 25: 401–414.

Seidenfeld, T., J. Kadane, and M. Schervish (1989), "On the shared preferences of two Bayesian decision makers", *J. Phil.* 86: 225–244.

Seidenfeld, T., M. Schervish, and J. Kadane (1995), "A representation of partially ordered preferences", *Ann. Statist.* 23: 2168–2217.

Seidenfeld, T. and L. Wasserman (1993), "Dilation for convex sets of probabilities", *Ann. Statist.* 21: 1139–1154.

Smith, C. A. B. (1961), "Consistency in statistical inference and decision", *J. Roy. Statist. Soc. B.* 23: 1–25.

Walley, P. (1991), *Statistical Reasoning With Imprecise Probabilities.* London: Chapman and Hall (Monographs on Statistics and Applied Probability).

———. (1996), "Inference from multinomial data: learning from a bag of marbles" (with discussion), *J. Roy. Statist. Soc. B.* 58: 3–58.

Walley, P. and T. L. Fine (1982), "Towards a frequentist theory of upper and lower probabilities", *Ann. Statist.* 10: 741–761.

Wasserman, L. A. and J. Kadane (1992), "Symmetric upper probabilities", *Ann. Statist.* 20: 1720–1736.