The Annals of Statistics
1993, Vol. 21, No. 3, 1139–1154

DILATION FOR SETS OF PROBABILITIES

BY TEDDY SEIDENFELD[1] AND LARRY WASSERMAN[2]

Carnegie Mellon University

Suppose that a probability measure P is known to lie in a set of probability measures M. Upper and lower bounds on the probability of any event may then be computed. Sometimes, the bounds on the probability of an event A conditional on an event B may strictly contain the bounds on the unconditional probability of A. Surprisingly, this might happen for every B in a partition $\mathscr{B}$. If so, we say that dilation has occurred. In addition to being an interesting statistical curiosity, this counterintuitive phenomenon has important implications in robust Bayesian inference and in the theory of upper and lower probabilities. We investigate conditions under which dilation occurs and we study some of its implications. We characterize dilation immune neighborhoods of the uniform measure.

1. Introduction. If M is a set of probability measures, then $\overline{P}(A) = \sup_{P \in M} P(A)$ and $\underline{P}(A) = \inf_{P \in M} P(A)$ are called the upper and lower probability of A, respectively. Upper and lower probabilities have become increasingly more common for several reasons. First, they 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) and Wasserman and Kadane (1992)]. Second, they arise in group decision problems [Levi (1982) and Seidenfeld, Schervish and Kadane (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), Levi (1974), Seidenfeld, Schervish and Kadane (1990) and Walley (1991)]. Fourth, sets of probabilities may result from incomplete or partial elicitation. Finally, there is some evidence that certain physical phenomena might be described by upper and lower probabilities [Fine (1988), and Walley and Fine (1982)].

Good (1966, 1974), in response to comments by Levi and Seidenfeld, Seidenfeld (1981) and Walley (1991) all have pointed out that it may sometimes happen that the interval $[\underline{P}(A), \overline{P}(A)]$ is strictly contained in the interval $[\underline{P}(A|B), \overline{P}(A|B)]$ for every B in a partition $\mathscr{B}$. In this case, we say that $\mathscr{B}$ dilates A. It is not surprising that this might happen for some B. What is surprising, is that this can happen no matter what $B \in \mathscr{B}$ occurs. Consider the following example [Walley (1991), pages 298–299].

Received October 1991; revised September 1992.
[1]Supported by NSF Grant SES-92-089428.
[2]Supported by NSF Grant DMS-90-05858.
AMS 1991 *subject classifications*. Primary 62F15; secondary 62F35.
Key words and phrases. Conditional probability, density ratio neighborhoods, ε-contaminated neighborhoods, robust Bayesian inference, upper and lower probabilities.

Suppose we flip a fair coin twice but the flips may not be independent. Let H_i refer to heads on toss i and T_i tails on toss i, $i = 1, 2$. Let M be the set of all P such that $P(H_1) = P(H_2) = \frac{1}{2}$, and $P(H_1 \text{ and } H_2) = p$ where $0 \le p \le \frac{1}{2}$. Now suppose we flip the coin. Then $P(H_2) = \frac{1}{2}$ but

$$0 = \underline{P}(H_2|H_1) < \underline{P}(H_2) = \tfrac{1}{2} = \overline{P}(H_2) < \overline{P}(H_2|H_1) = 1$$

and

$$0 = \underline{P}(H_2|T_1) < \underline{P}(H_2) = \tfrac{1}{2} = \overline{P}(H_2) < \overline{P}(H_2|T_1) = 1.$$

We begin with precise beliefs about the second toss and then, no matter what happens on the first toss, merely learning that the first toss has occurred causes our beliefs about the second toss to become completely vacuous. The important point is that this phenomenon occurs *no matter what the outcome of the first toss was*. This goes against our seeming intuition that when we condition on new evidence, upper and lower probabilities should shrink toward each other.

Dilation leads to some interesting questions. For example, suppose the coin is tossed and we observe the outcome. Are we entitled to retain the more precise unconditional probability instead of conditioning? See Levi (1977) and Kyburg (1977) for discussion on this.

To emphasize the counterintuitive nature of dilation, imagine that a physician tells you that you have probability $\frac{1}{2}$ that you have a fatal disease. He then informs you that he will carry out a blood test tomorrow. Regardless of the outcome of the test, if he conditions on the new evidence, he will then have lower probability 0 and upper probability 1 that you have the disease. Should you allow the test to be performed? Is it rational to pay a fee not to perform the test?

The behavior is reminiscent of the nonconglomerability of finitely additive probabilities. For example, if P is finitely additive, there may be an event A and a partition $\mathscr{B}$ such that $P(A) = \frac{1}{4}$, say, but $P(A|B_i) = \frac{3}{4}$ for every $B_i \in \mathscr{B}$. See Schervish, Seidenfeld and Kadane (1984). A key difference, however, is that nonconglomerability involves infinite spaces whereas dilation occurs even on finite sets—dilation cannot be explained as a failure of our intuition on infinite sets. A key similarity is that both phenomena entail a difference between decisions in normal and extensive form [Seidenfeld (1991)]. It is interesting to note that Walley (1991) regards nonconglomerability as incoherent but he tolerates dilation.

The purpose of this paper is to study the phenomenon of dilation and to investigate its ramifications. We believe that this is the first systematic study of dilation. As we shall point out, dilation has implications for elicitation, robust Bayesian inference and the theory of upper and lower probabilities. Furthermore, we will show that dilation is not a pathological phenomenon.

In Section 2 we define dilation and we give some characterizations of its occurrence. Examples are studied in Section 3 with particular emphasis on ε-contaminated models. Section 4 characterizes dilation immune neighborhoods. Finally, we discuss the results in Section 5.

2. Dilation. Let Ω be a nonempty set and let $\mathscr{E}(\Omega)$ be an algebra of subsets of Ω. Let $\mathscr{P}$ be the set of all probability measures on $\mathscr{E}(\Omega)$. If M is a set of probability measures on $\mathscr{E}(\Omega)$, define the *upper probability function* $\overline{P}$ and *lower probability function* $\underline{P}$ by

$$\underline{P}(A) = \inf_{P \in M} P(A) \quad \text{and} \quad \overline{P}(A) = \sup_{P \in M} P(A).$$

We assume that M is convex and closed with respect to the total variation topology.

The most common situation where a set of probabilities M would arise is in the theory of robustness. In classical robustness [Huber (1981, 1973)] M is a class of sampling models. In Bayesian robustness [Berger (1984) and Lavine (1991)] M is a set of prior distributions. Another way the sets of probabilities arise is through the theory of upper and lower probabilities. For example, Smith (1961) and Walley (1991), among others, show that if the axioms of probability are weakened, then we end up with upper and lower probabilities. This approach is a generalization of de Finetti's (1964) notion of coherence.

If $\underline{P}(B) > 0$, define the *conditional upper and lower probability given* B by

$$\underline{P}(A|B) = \inf_{P \in M} P(A \cap B)/P(B) \quad \text{and} \quad \overline{P}(A|B) = \sup_{P \in M} P(A \cap B)/P(B).$$

This is the natural way to define an upper and lower conditional probability if the robustness point of view is taken. If we follow the axiomatic approach of Walley (1991), then this way of defining upper and lower conditional probabilities can be justified through a coherence argument.

Say that B *dilates* A and write $B \rightsquigarrow A$ if $[\underline{P}(A), \overline{P}(A)]$ is strictly contained in $[\underline{P}(A|B), \overline{P}(A|B)]$. (Here we mean strict containment in the set-theoretic sense.) If $\mathscr{B}$ is a finite partition for which $\underline{P}(B) > 0$ for all $B \in \mathscr{B}$, then we say that $\mathscr{B}$ *dilates* A and we write $\mathscr{B} \rightsquigarrow A$ if $B \rightsquigarrow A$ for every $B \in \mathscr{B}$. We will say that M is *dilation prone* if there exists A and $\mathscr{B}$ such that $\mathscr{B} \rightsquigarrow A$. Otherwise, M is *dilation immune*. We will say that $\mathscr{B}$ *strictly dilates* A if $\underline{P}(A|B) < \underline{P}(A) \le \overline{P}(A) < \overline{P}(A|B)$ for every $B \in \mathscr{B}$. Obviously, if either $B \subset A$ or $B \subset A^c$ for some $B \in \mathscr{B}$, then dilation is impossible. Hence, we shall assume that $A \cap B \neq \varnothing$ and $A^c \cap B \neq \varnothing$ for all $B \in \mathscr{B}$.

Given M, define

$$M_*(A) = \{P \in M;\, P(A) = \underline{P}(A)\}, \qquad M^*(A) = \{P \in M;\, P(A) = \overline{P}(A)\}$$

and

$$M_*(A|B) = \{P \in M;\, P(A|B) = \underline{P}(A|B)\},$$

$$M^*(A|B) = \{P \in M;\, P(A|B) = \overline{P}(A|B)\}.$$

It will be useful to define the following two notions of dependence. For $P \in \mathscr{P}$ define $S_P(A, B) = P(A \cap B)/(P(A)P(B))$ if $P(A)P(B) > 0$ and $S_P(A, B) = 1$ if $P(A)P(B) = 0$ and also define $d_P(A, B) = P(A \cap B) - P(A)P(B)$. Note that $S_P(A, B) < 1$ if and only if $S_P(A, B^c) > 1$. (This is a consequence of

Lemma 3.2.) Also define

$$\Sigma^{+}(A,B) = \{P \in \mathscr{P}; d_P(A,B) > 0\} \quad \text{and}$$
$$\Sigma^{-}(A,B) = \{P \in \mathscr{P}; d_P(A,B) < 0\}.$$

The surface of independence for events A and B is defined by

$$\mathscr{I}(A,B) = \{P \in \mathscr{P}; d_P(A,B) = 0\}.$$

The next four theorems show that the independence surface plays a crucial role in dilation. A necessary condition for dilation is that the independence surface cuts through M (Theorem 2.1). But this condition is not sufficient. A sufficient condition is given in Theorem 2.3. A variety of cases exist in between. These are explored in Sections 3 and 4.

THEOREM 2.1. *Let $\mathscr{B} = \{B, B^c\}$. If $\mathscr{B}$ dilates A, then $M \cap \mathscr{I}(A,B) \neq \emptyset$.*

PROOF. Choose $P \in M_*(A|B)$. Then $\underline{P}(A|B) = P(A|B) = S_P(A,B)P(A)$. Dilation implies that $S_P(A,B)P(A) \leq \underline{P}(A)$. Thus, $S_P(A,B) \leq \underline{P}(A)/P(A) \leq 1$. Similarly, there exists $Q \in M$ such that $S_Q(A,B) \geq 1$. Let $R_\alpha = \alpha P + (1-\alpha)Q$ and let $S_\alpha = S_{R_\alpha}(A,B)$. Then S_α is a continuous function of α and by the intermediate value theorem, there is an $a \in (0,1)$ such that $S_a = 1$. Thus, $R_a \in \mathscr{I}(A,B)$ and by convexity of M $R_a \in M$. □

If $\mathscr{B}$ strictly dilates A, then M need not be closed for the previous result.

THEOREM 2.2. *If $\mathscr{B}$ strictly dilates A, then for every $B \in \mathscr{B}$, $M_*(A|B) \subset \Sigma^-(A,B)$ and $M^*(A|B) \subset \Sigma^+(A,B)$.*

PROOF. Choose $P \in M_*(A|B)$. Then $P(AB)/P(B) = \underline{P}(A|B) < \underline{P}(A) \leq P(A)$. Hence, $S_P(A,B) < 1$ so that $P \in \Sigma^-(A,B)$. Similarly for the other case. □

THEOREM 2.3. *If for every $B \in \mathscr{B}$,*

$$M_*(A) \cap \Sigma^-(A,B) \neq \emptyset \quad \textit{and} \quad M^*(A) \cap \Sigma^+(A,B) \neq \emptyset,$$

then $\mathscr{B}$ strictly dilates A.

PROOF. Choose $P \in M_*(A) \cap \Sigma^-(A,B)$. Then $P(A) = \underline{P}(A)$ and $P(AB) < P(A)P(B)$. Thus, $\underline{P}(A) = P(A) > P(AB)/P(B) = P(A|B) \geq \underline{P}(A|B)$. A similar argument applies for the upper bound. □

Many axiomatic approaches to probability involve an assumption that we can enlarge the space and include events with given probabilities. For example, we might assume that we can add an event that corresponds to the flip of a coin with a prescribed probability p. DeGroot (1970), Koopman (1940) and Savage (1972) all make an assumption of this nature. If we include this assumption, then as the next theorem shows, dilation always occurs with nontrivial upper and lower probabilities. For any event A, define $\mathscr{A}(A) = \{\varnothing, A, A^c, A \cup A^c\}$. Given two algebras $\mathscr{B}$ and $\mathscr{C}$, let $\mathscr{B} \otimes \mathscr{C}$ be the product algebra.

THEOREM 2.4. *Suppose there exists* $E \in \mathscr{B}$ *such that* $0 < \alpha < \beta < 1$, *where* $\alpha = \underline{P}(E)$ *and* $\beta = \overline{P}(E)$. *Let* A *and* B *be events and let* $\mathscr{B}' = \mathscr{B} \otimes \mathscr{A}(A) \otimes \mathscr{A}(B)$. *For* $\lambda \in [0,1]$ *define a set of probabilities on* $\mathscr{B}'$ *by* $M' = \{P \otimes P_{0.5} \otimes P_\lambda;\ P \in M\}$, *where* $P_{0.5}(A) = \frac{1}{2}$, $P_\lambda(B) = \lambda$. *Here* $P \otimes P_{0.5} \otimes P_\lambda$ *is the product measure on* $\mathscr{B}'$ *with* P, $P_{0.5}$ *and* P_λ *as marginals. Then there exists* λ *such that strict dilation occurs in* M'.

PROOF. We prove the case where $\beta < \frac{1}{2}$. Choose λ such that

$$\frac{0.5 - \beta}{1 - \beta} < \lambda < \frac{0.5 - \alpha}{1 - \alpha}.$$

Let $F = (\Omega \times A \times B) \cup (E \times A \times B^c) \cup (E^c \times A^c \times B^c)$. Then $\underline{P}(F) = \overline{P}(F) = \frac{1}{2}$ and $\underline{P}(F|A) = \lambda + (1 - \lambda)\alpha < \frac{1}{2} < \lambda + (1 - \lambda)\beta = \overline{P}(F|A)$. Also, $\underline{P}(F|A^c) = (1 - \lambda)(1 - \beta) < \frac{1}{2} < (1 - \lambda)(1 - \alpha) = \overline{P}(F|A^c)$. Thus, $\{A, A^c\}$ strictly dilates F. $\square$

REMARK. Theorems 2.2 through 2.4 are still true even if the convexity of M is dropped. Also, Theorem 2.4 does not require closure.

3. Examples. In this section we consider classes of probabilities that are common in Bayesian robustness [Berger (1984, 1985, 1990)] and we find conditions for dilation. A detailed investigation of a certain class of upper probabilities is given in Section 4.

EXAMPLE 3.1. In between the necessary condition of Theorem 2.3 and the sufficient condition of Theorem 2.4 are many cases. This is illustrated with the following example. Let $P \in \Sigma^-(A, B)$ and $Q \in \Sigma^+(A, B)$. Let M be the convex hull of P and Q. Thus, M is a line segment. If $P(A) = Q(A)$, then there is dilation. In other words, if the line segment M is parallel to the side of the simplex corresponding to the event A, there is dilation. This is the sufficient condition of Theorem 2.4. Now suppose that $P(A) \le Q(A)$ and define the angle of Q with respect to P by

$$\text{angle}(Q, P) = \exp\left\{\left|\log \frac{Q(A)}{P(A)}\right| - |\log S_Q|\right\}$$

and the angle of P with respect to Q by

$$\text{angle}(P, Q) = \exp\left\{\left|\log \frac{P(A)}{Q(A)}\right| - |\log S_P|\right\}.$$

Then there is dilation if and only if both angles are less than 1. In other words, dilation occurs when the line segment is sufficiently "perpendicular" to the surface of independence.

EXAMPLE 3.2 (ε-contaminated classes). The most common class of probabilities that are used in robustness is the ε-contaminated class [Huber (1973, 1981) and Berger (1984, 1985, 1990)] that is defined by $M = \{(1-\varepsilon)P + \varepsilon Q;\ Q \in \mathscr{P}\}$, where P is a fixed probability measure and ε is a fixed number in $[0, 1]$. To avoid triviality, assume $\varepsilon > 0$ and that P is an internal point in the set of all probability measures.

LEMMA 3.1. *Dilation occurs for this class if and only if*

$$\varepsilon > \max\left\{\frac{d_P(A,B)}{P(A)P(B^c)}, \frac{d_P(A,B)}{P(A^c)P(B)}, -\frac{d_P(A,B)}{P(A^c)P(B^c)}, -\frac{d_P(A,B)}{P(A)P(B)}\right\}.$$

PROOF. Note that

$$\underline{P}(A) = (1-\varepsilon)P(A) \quad \text{and} \quad \underline{P}(A|B) = \frac{(1-\varepsilon)P(AB)}{(1-\varepsilon)P(B)+\varepsilon}.$$

Thus, $\underline{P}(A|B) < \underline{P}(A)$ if and only if $d_P(A,B) < \varepsilon P(A)P(B^c)$. The other inequalities follow from similar computations. □

REMARK. If $d_P(A, B) = 0$, then dilation occurs for every $\varepsilon > 0$.

It is useful to reexpress the above result as follows. Using that fact that $P(B)(S_P(A^c, B) - 1) = P(B^c)(1 - S_P(A^c, B^c))$ and using the fact that $S_P(\cdot\,, \cdot)$ is symmetric in its arguments, $P(A)(1 - S_P(A, B)) = P(A^c)(S_P(A^c, B) - 1)$, so we have (with obvious generalization to larger partitions):

$$\underline{P}(A|B) < \underline{P}(A) \quad \text{if and only if } \varepsilon > (1 - S_P(A, B^c))$$

and

$$\overline{P}(A|B) > \overline{P}(A) \quad \text{if and only if } \varepsilon > (1 - S_P(A^c, B^c)).$$

If P is a nonatomic measure on the real line, then there always exist A and B with positive probability that are independent under P. Thus, $S_P = 1$ and hence dilation occurs for every $\varepsilon > 0$.

To pursue this example further, we now investigate the behavior of dilation over subpartitions. Specifically, we show that if there is a partition that strictly dilates A, then there is a binary partition that strictly dilates A. To prove this, we need a few lemmas that apply generally. The proofs of the next three lemmas are straightforward and are omitted.

LEMMA 3.2. $\sum_{i=1}^{n} S_P(A, B_i)P(B_i) = 1$.

LEMMA 3.3. *Let $B = C \cup D$ for events $C \cap D = \varnothing$. Then*

$$S_P(A, B) = \frac{P(C)S_P(A, C) + P(D)S_P(A, D)}{P(B)}.$$

REMARK. Note that the lemma generalizes in an obvious way for a finite set of disjoint events.

REMARK. If $\pi = \{C_1, \dots, C_n\}$ dilates A, then π dilates A^c.

LEMMA 3.4. *ε-contamination is preserved under subalgebras. That is, if $M = \{(1-\varepsilon)P + \varepsilon Q;\ Q \text{ arbitrary}\}$ and $\mathscr{A}$ is a subalgebra of $\mathscr{E}(\Omega)$, then $M_{\mathscr{A}} = \{(1-\varepsilon)P_{\mathscr{A}} + \varepsilon Q;\ Q \text{ arbitrary}\}$, where $M_{\mathscr{A}} = \{P_{\mathscr{A}};\ P \in M\}$ and $P_{\mathscr{A}}$ is the restriction of P to $\mathscr{A}$.*

THEOREM 3.1. *Let M be an ε-contaminated class. Suppose that $\pi_n = \{C_1, \dots, C_n\}$ is a finite partition that strictly dilates A. Then there exists a binary partition $\mathscr{B} = \{B, B^c\}$ that strictly dilates A.*

PROOF. Assume that $n \geq 3$ and there is no strict dilation in any coarser partition $\pi' \subset \pi_n$. We have that

$$\underline{P}(A|C_i) < \underline{P}(A) \leq \overline{P}(A) < \overline{P}(A|C_i).$$

Define three families of events from π_n by

$$S^+ = \{E;\ S_P(A, E) > 1\}, \qquad S^- = \{E;\ S_P(A, E) < 1\},$$
$$S^1 = \{E;\ S_P(A, E) = 1\}.$$

By Lemma 3.1 we know that independence is sufficient for dilation in an ε-contamination model. Hence, if $S^1 \neq \varnothing$, we are done. So assume $S^1 = \varnothing$.

Let $C^+ = \{C_i \in \pi_n;\ C_i \in S^+\}$ and $(C^+)^c = C^- = \{C_i \in \pi_n;\ C_i \in S^-\}$. From the assumption that π_n strictly dilates A, by the remark following Lemma 3.1,

$$\varepsilon > (1 - S_P(A, C_i^c)) \quad \text{if } C_i \in C^+ \tag{1}$$

and

$$\varepsilon > (1 - S_P(A^c, C_i^c)) \quad \text{if } C_i \in C^-. \tag{2}$$

From Lemma 3.2 and 3.4 and the assumption that there is no strict dilation in the partition $\{C_i, C_i^c\}$, we conclude that

$$\varepsilon \leq (1 - S_P(A^c, C_i)) \quad \text{if } C_i \in C^+ \tag{3}$$

and

$$\varepsilon \leq (1 - S_P(A, C_i)) \quad \text{if } C_i \in C^-. \tag{4}$$

Let k_+ be the cardinality of C^+ and let k_- be the cardinality of C^-. Without loss of generality, assume that $k = k_+ \leq k_-$. Write $\pi_n = \{C_1, \dots, C_k, C_{k+1}, \dots, C_n\}$ where $C^+ = \{C_1, \dots, C_k\}$. Consider the events $E_{ij} = (C^+ - C_i) \cup (C^- - C_j)$ for $C_i \in C^+$ and $C_j \in C^-$.

Case 1: If there exists i, j such that $S_P(A, E_{ij}) < 1$, then by Lemmas 3.2 and 3.4, $S_P(A, (C_i \cup C_j)) > 1$ since $E_{ij}^c = (C_i \cup C_j)$. We arrive at the following contradiction. Because there is no dilation in $\pi_{ij} = \{C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_{j-1}, C_{j+1}, \dots, (C_i \cup C_j)\}$, by (1) we know that

$$(*) \qquad \varepsilon \leq \left(1 - S_P(A, E_{ij})\right).$$

Since $C_j \in C^-$ and since $\{C_j, C_j^c\}$ does not dilate A, by (4),

$$(**) \qquad \varepsilon \leq \left(1 - S_P(A, C_j)\right).$$

But by assumption of dilation in π_n, we have, from (1),

$$(***) \qquad \varepsilon > \left(1 - S_P(A, C_i^c)\right).$$

However, $C_i^c = (E_{ij} \cup C_j)$. Hence, by Lemma 3.3, using $(*)$ and $(**)$,

$$\varepsilon \leq 1 - S_P(A, E_{ij} \cup C_j) = 1 - S_P(A, C_i^c),$$

which contradicts $(***)$.

Case 2: We have that $S_P(A, E_{ij}) > 1$ for all i, j so that by Lemma 3.2, $S_P(A, E_{ij}^c) < 1$. Thus, as $E_{ij}^c = (C_i \cup C_j)$, $S_P(A, (C_i \cup C_j)) < 1$ for all $C_i \in C^+$, $C_j \in C^-$. Since $k_+ \leq k_-$ we may form the k disjoint pairs $F_1 = C_1 \cup C_{k+1}, \dots, F_k = C_k \cup C_{k+k}$ and $S_P(A, F_i) < 1$, $i = 1, \dots, k$. Let $F = \bigcup_i F_i$. By Lemma 3.3, $S_P(A, F) < 1$. However, $C^+ \subset F$ so either F^c is empty or $F^c \subset S^-$ so that $S_P(A, F) \geq 1$, a contradiction. $\square$

The following example illustrates how the dilation preserving coarsenings may be quite limited. Let $P(AC_1) = \frac{1}{4}$, $P(AC_2) = \frac{1}{6}$, $P(AC_3) = \frac{1}{12}$, $P(A^cC_1) = \frac{1}{12}$, $P(A^cC_2) = \frac{1}{6}$ and $P(A^cC_3) = \frac{1}{4}$. So $P(A) = P(A^c) = \frac{1}{2}$ and $P(C_i) = \frac{1}{3}$ for $i = 1, 2, 3$. Note that $S_P(A, C_i) = \frac{3}{2}, 1, \frac{1}{2}$ for $i = 1, 2, 3$. If $\varepsilon > \frac{1}{4}$, then $\{C_1, C_2, C_3\}$ dilates A. Since $S_P(A, C_2) = 1$, $\{C_2, C_2^c\}$ dilates A for every $\varepsilon > 0$. However, if $\varepsilon < \frac{1}{2}$, neither $\{C_1, C_1^c\}$ nor $\{C_3, C_3^c\}$ dilates A.

Also, the ε-contaminated model has the property that the upper and lower conditionals cannot shrink inside $\underline{P}(A)$ and $\overline{P}(A)$. Specifically, note that $\underline{P}(A|B) < \underline{P}(A)$ if and only if $\varepsilon > (1 - S_P(A, B^c))$ and $\overline{P}(A|B) > \overline{P}(A)$ if and only if $\varepsilon > (1 - S_P(A^c, B^c))$. At least one of $S_P(A, B^c)$ and $S_P(A^c, B^c)$ must be greater than or equal to 1 so that at least one of these inequalities must occur. Hence, it cannot be that $\underline{P}(A) \leq \underline{P}(A|B) \leq \overline{P}(A|B) \leq \overline{P}(A)$.

EXAMPLE 3.3 (Total variation neighborhoods). Define the total variation metric by $\delta(P, Q) = \sup_A |P(A) - Q(A)|$. Fix P and ε and assume that P is internal. Let $M = \{Q; \delta(P,Q) \leq \varepsilon\}$. Then $\underline{P}(A) = \max\{P(A) - \varepsilon, 0\}$ and $\overline{P}(A) = \min\{P(A) + \varepsilon, 1\}$. Also,

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

There are four cases:

Case 1: $P(AB)$, $P(AB^c) \leq \varepsilon$. Dilation occurs if and only if

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

Case 2: $P(AB) \leq \varepsilon < P(AB^c)$. Dilation occurs if and only if

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

Case 3: $P(AB^c) \leq \varepsilon < P(AB)$. Dilation occurs if and only if

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

Case 4: $\varepsilon < P(AB)$, $P(AB^c)$. Dilation occurs if and only if

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

EXAMPLE 3.4 (Density ratio classes). Let $\Omega = \{\omega_1, \dots, \omega_n\}$ and let $p = (p_1, \dots, p_n)$ be a probability vector with each $p_i > 0$. For $k \geq 1$, define the density ratio neighborhood by $M_k = \{q = (q_1, \dots, q_n);\ q_i/q_j \leq kp_i/p_j$ for all $i,j\}$. [A more general case is considered in DeRobertis and Hartigan (1981). Also, see Section 4 of this paper.] Then $\underline{P}(A) = P(A)/(P(A) + kP(A^c))$, where P is the probability measure generated by p. Also, $\underline{P}(A|B) = P(AB)/(P(AB) + kP(A^cB))$. If $d_P(A,B) = 0$, then $[\underline{P}(A|B), \overline{P}(A|B)] = [\underline{P}(A), \overline{P}(A)]$ so dilation does not occur. If $d_P(A,B) > 0$, then $\underline{P}(A|B) > \underline{P}(A)$ so dilation does not occur. If $d_P(A,B) < 0$, then $\underline{P}(A|B^c) > \underline{P}(A)$ so dilation does not occur. Thus, dilation never occurs. This class also posseses many other interesting properties—see Wasserman (1992).

4. Neighborhoods of the uniform measure. In Bayesian robustness it is common to use sets of probabilities that are neighborhoods of a given probability measure. In this section we investigate neighborhoods of the uniform measure on a compact set. Subject to some mild regularity conditions, we characterize dilation immune neighborhoods. Specifically, we show that the only dilation immune neighborhoods are the density ratio neighborhoods. This has important implications in Bayesian robustness since it means that, unless these neighborhoods are used, dilation will be the rule rather than the exception. It also shows that it is the structure of the class, not necessarily its size, that causes dilation. The mathematical techniques used here are based on continuous majorization theory as developed in Ryff (1965). See also Hardy, Littlewood and Pólya (1952), Chapter 10, and Marshall and Olkin (1979). The restriction to neighborhoods around the uniform is, of course, a special case. But the restriction to this special case allows for an intense investigation of the phenomenon. Furthermore, neighborhoods of uniform priors are an important special case in Bayesian robustness.

Let $\Omega = [0,1]$, let $\mathscr{E}(\Omega)$ be the Borel sets and let μ be Lebesgue measure. Given two measurable functions f and g, say that f and g are *equimeasurable* and write $f \sim g$ if $\mu(\{\omega;\ f(\omega) > t\}) = \mu(\{\omega;\ g(\omega) > t\})$ for all t. Loosely speaking, this means that g is a "permutation" of f. Given f, there is a unique, nonincreasing, right-continuous function f^* such that $f^* \sim f$. The function f^* is called the *decreasing rearrangement* of f. We say that f is *majorized* by g and we write $f \prec g$ if $\int_0^1 f = \int_0^1 g$ and $\int_0^s f^* \le \int_0^s g^*$ for all s. Here $\int f$ means $\int f(\omega)\mu(d\omega)$. Let $\Lambda(f)$ be the convex closure of $\{g;\ g \sim f\}$. Ryff (1965) shows that $\Lambda(f) = \{g;\ g \prec f\}$. We define the *increasing rearrangement* of f to be the unique, nondecreasing, right-continuous function f_* such that $f_* \sim f$.

Let $u(\omega) = 1$ for all $\omega \in \Omega$. Let m be a weakly closed, convex set of bounded density functions with respect to Lebesgue measure on Ω, let M be the corresponding set of probability measures and let $\overline{P}$ and $\underline{P}$ be the upper and lower probability generated by M. We call m a neighborhood of u if $f \in m$ implies that $g \in m$ whenever $g \sim f$. This condition is like requiring permutation invariance for neighborhoods of the uniform measure on finite sets. All common neighborhoods satisfy this regularity condition. From Ryff's theorem, it follows that if $f \in m$ and $g \prec f$, then $g \in m$. The properties of such sets are studied in Wasserman and Kadane (1992). If m is a neighborhood of u, we shall say that M is a neighborhood of μ. To avoid triviality, we assume that $M \neq \{\mu\}$. Next, we state a useful lemma. The proof is by direct calculation and is omitted.

LEMMA 4.1. *Let $\mathscr{B} = \{B_1, \dots, B_n\}$ be a finite partition of Ω and let f and g be two probability density functions such that $\int_{B_i} f = \int_{B_i} g$ for $i = 1, \dots, n$. If f is constant over each B_i, then $f \prec g$.*

For every f define

$$\rho(f) = \frac{\operatorname{ess\,sup} f}{\operatorname{ess\,inf} f},$$

where $\operatorname{ess\,sup} f = \inf\{t;\ \mu(\{\omega;\ f(\omega) > t\}) = 0\}$ and $\operatorname{ess\,inf} f = \sup\{t;\ \mu(\{\omega;\ f(\omega) < t\}) = 0\}$. For $k \ge 1$, define $\gamma_k = \{f;\ \rho(f) \le k\}$. This is the *density ratio neighborhood of* μ. Let $\mathscr{A} = \{A \in \mathscr{E}(\Omega);\ 0 < \mu(A) < 1\}$. For every $A \in \mathscr{A}$ define a density f_A by

$$f_A(\omega) = \begin{cases} \dfrac{\underline{P}(A)}{\mu(A)}, & \text{if } \omega \in A, \\[2ex] \dfrac{1 - \underline{P}(A)}{1 - \mu(A)}, & \text{if } \omega \in A^c. \end{cases}$$

Let $P_A(d\omega) = f_A(\omega)\mu(d\omega)$. Note that $P_A(A) = \underline{P}(A)$. Let $m_*(A) = \{dP/d\mu;$

$P \in M_*(A)\}$. Also define f^A by

$$f^A(\omega) = \begin{cases} \dfrac{\overline{P}(A)}{\mu(A)}, & \text{if } \omega \in A, \\ \dfrac{1 - \overline{P}(A)}{1 - \mu(A)}, & \text{if } \omega \in A^c. \end{cases}$$

We define P^A and $m^*(A)$ analogously.

LEMMA 4.2. *Suppose that m is a neighborhood of u. For every $A \in \mathscr{A}$, $f_A \in m_*(A)$.*

PROOF. Choose $f \in m_*(A)$. Then by Lemma 4.1, $f_A \prec f$ so $f_A \in m$. But $\int_A f_A = \underline{P}(A)$ so $f_A \in m_*(A)$. □

For every $t \in (0,1)$ let $A_t = [0,t]$, $m_t = m_*(A_t)$, $m^t = m^*(A_t)$, $f^t = f^{A_t}$ and $f_t = f_{A_t}$. Define $\overline{F}(t) = \overline{P}(A_t)$ and $\underline{F}(t) = \underline{P}(A_t)$. Also, define

$$k^t = \left(\frac{\overline{F}(t)}{1 - \overline{F}(t)}\right)\left(\frac{1-t}{t}\right) \quad \text{and} \quad k_t = \left(\frac{1 - \underline{F}(t)}{\underline{F}(t)}\right)\left(\frac{t}{1-t}\right).$$

Let $c^t = 1/(k^t t + (1-t))$ and $c_t = 1/(t + k_t(1-t))$. Then f^t is equal to $c^t k^t$ on A_t and is equal to c^t on A_t^c. Similarly, f_t is equal to c_t on A_t and is equal to $c_t k_t$ on A_t^c. It is easy to show that $k^t \ge 1$ and $k_t \ge 1$. If $\mu(A) = \mu(B)$, then $\underline{P}(A) = \underline{P}(B)$ and $\overline{P}(A) = \overline{P}(B)$. Also, $\underline{F}(t) + \overline{F}(1-t) = 1$. Hence, $k_t = k^{1-t}$. In particular, $k^{1/2} = k_{1/2} = \hat{k}$, say.

LEMMA 4.3. *Suppose that M is a dilation immune neighborhood of μ. Then, for all $t \in (0,1)$, $k^t \le \hat{k}$ and $k_t \le \hat{k}$.*

PROOF. Consider $t \in (0, \frac{1}{2})$. Suppose that $k^t > \hat{k}$. There exists $n \ge 1$ such that $nt \le \frac{1}{2} < (n+1)t$. Suppose that $nt < \frac{1}{2}$—the proof for the case where $nt = \frac{1}{2}$ is similar. Define $W_i = [(i-1)t, it)$ for $i = 1, \dots, n$ and $W_{n+1} = [nt, \frac{1}{2})$. Define $Y_i = [\frac{1}{2} + (i-1)t, \frac{1}{2} + it)$ for $i = 1, \dots, n$ and $Y_{n+1} = [\frac{1}{2} + nt, 1]$. Let $B_i = W_i \cup Y_i$, $A = A_{1/2}$ and $P(d\omega) = f^{1/2}(\omega)\mu(d\omega)$. Then A is independent of each B_i under P. Let f_i be a rearrangement of f^t that is equal to $c^t k^t$ on W_i and that is equal to c^t over all of $A_{1/2}^c$. [This is possible since each W_i has $\mu(W_i) \le t$.] Let $Q_i(d\omega) = f_i(\omega)\mu(d\omega)$ and $a_i = \mu(W_i) = \mu(Y_i)$. Then

$$\begin{aligned} \overline{P}(A|B_i) \ge Q_i(A|B_i) &= \frac{Q_i(W_i)}{Q_i(W_i) + Q_i(Y_i)} \\ &= \frac{1}{1 + (Q_i(Y_i)/Q_i(W_i))} = \frac{1}{1 + (c^t a_i / c^t k^t a_i)} \\ &= \frac{1}{1 + (1/k^t)} > \frac{1}{1 + (1/\hat{k})} \\ &= P(A|B_i) = P(A) = \overline{P}(A). \end{aligned}$$

Now, $\underline{P}(A|B_i) \leq P_A(A|B_i) = P_A(A) = \underline{P}(A)$. Hence, $\{B_1, \ldots, B_{n+1}\} \rightsquigarrow A$ which is a contradiction. Thus, $k^t \leq \hat{k}$. A similar argument shows that $k_t \leq \hat{k}$. The relation $k_t = k^{1-t}$ establishes the result for all $t \in (0,1)$. □

LEMMA 4.4. *Suppose that M is a dilation immune neighborhood of μ. Then $k^t = k_t = \hat{k}$ for all $t \in (0,1)$.*

PROOF. We prove the result for k^t. Consider $t \in (0, \frac{1}{2})$. Suppose that $k^t < \hat{k}$. Let $P(d\omega) = f^{1/2}(\omega)\mu(d\omega)$, $Q(d\omega) = f^t(\omega)\mu(d\omega)$ and $R(d\omega) = f_t(\omega)\mu(d\omega)$. Let $W = [0, \omega]$ and $Y = [\frac{1}{2}, 1]$ where $w = t/(2(1-t))$. Note that $w < t$ so that $W \subset A_t$. Let $B = W \cup Y$ and $A = A_t$. Then A and B are independent under both Q and R. By a similar argument as that in the previous lemma, we deduce from the fact that $k^t < \hat{k}$, that $\overline{P}(A|B) \geq P(A|B) > Q(A|B) = Q(A) = \overline{P}(A)$. Let $\hat{f}$ equal $c^{1/2}\hat{k}$ on $[0, t] \cup (t + \frac{1}{2}, 1]$ and equal $c^{1/2}$ on $[t, t + \frac{1}{2}]$. Then $\hat{f} \sim f^{1/2}$ and $\overline{P}(A|B^c) \geq \hat{P}(A|B^c) > Q(A|B^c) = Q(A) = \overline{P}(A)$, where $\hat{P}(d\omega) = \hat{f}(\omega)\mu(d\omega)$. Also, $\underline{P}(A|B) \leq R(A|B) = R(A) = \underline{P}(A)$ and $\underline{P}(A|B^c) \leq R(A|B^c) = R(A) = \underline{P}(A)$. We have a dilation which is a contradiction; thus, $k^t \geq \hat{k}$. By a similar argument $k_t \geq \hat{k}$. From $k_t = k^{1-t}$ this holds for all $t \in (0,1)$. From the previous lemma, $k^t \leq \hat{k}$ and $k_t \leq \hat{k}$. Hence, the claim follows. □

Let $\overline{F}_k(t) = \overline{P}_k(A_t)$, where $\overline{P}_k$ is the upper probability generated by the density ratio neighborhood γ_k. The next lemma is a standard fact about density ratio neighborhoods and we state it without proof.

LEMMA 4.5. $\overline{F}_k(t) = kt/(kt + (1-t))$.

LEMMA 4.6. *If M is a dilation immune neighborhood of μ, then $\overline{F} = \overline{F}_k$ for some $k \geq 1$.*

PROOF. Follows from the last two lemmas. □

We conclude that if M is dilation immune, then M generates the same upper probability as γ_k. But this does not show that $m = \gamma_k$. To show this, we need one more lemma. Given $k \geq 1$ and $t \in (0,1)$, define $r^{k,t}$ by

$$r^{k,t}(\omega) = \begin{cases} c^{k,t}k, & \text{if } \omega \leq t, \\ c^{k,t}, & \text{if } \omega > t, \end{cases}$$

where $c^{k,t} = \{kt + (1-t)\}^{-1}$.

The next lemma gives an integral representation of density ratio neighborhoods.

LEMMA 4.7 (Integral representation of density ratio neighborhoods). *The following two statements are equivalent*:

(i) $f \in \gamma_k$.

(ii) *There exists a probability measure* R *on* $\mathscr{C}(\Omega)$ *and a number* $z \in [1, k]$ *such that for almost all* ω, $f^*(\omega) = \int_0^1 r^{z,t}(\omega) R(dt)$.

PROOF. Suppose (i) holds. Then $1 < z = f^*(0)/f^*(1) \leq k$. Let $h(\omega) = f^*(\omega)/f^*(1)$. Define a set function V by $V([0, \omega]) = (z - h(\omega))/(z - 1)$. Then $V(\Omega) = 1$ so V can be extended to be a probability measure on $\mathscr{C}(\Omega)$. Let $h^{z,t}(\omega) = r^{z,t}(\omega)/c^{z,t}$. Then, for almost all ω, $\int_0^1 h^{z,t}(\omega)V(dt) = V([0, \omega]) + z(1 - V([0, \omega])) = h(\omega)$. Now

$$f^*(\omega) = f^*(1)h(\omega) = f^*(1)\int_0^1 h^{z,t}(\omega)V(dt)$$
$$= \int_0^1 r^{z,t}(\omega)R(dt),$$

where R is defined by $R([0, \omega]) = V([0, \omega])f^*(1)/c^{z,t}$. Now we confirm that R is a probability measure. We have

$$1 = \int_0^1 f^*(\omega)\mu(d\omega) = \int_0^1\int_0^1 r^{z,t}(\omega)R(dt)\mu(d\omega)$$
$$= \int_0^1\int_0^1 r^{z,t}(\omega)\mu(d\omega)R(dt) = \int_0^1 R(dt) = R(\Omega).$$

Thus, R is a probability measure. Hence, (ii) holds.

Now suppose that (ii) holds. Then $\rho(f) = f^*(0)/f^*(1) = h(0) = z$ for some $z \in [1, k]$. Thus, (i) holds. □

THEOREM 4.1. *Suppose that* M *is a neighborhood of* μ. *Then* M *is dilation immune if and only if* $m = \gamma_k$ *for some* k.

PROOF. Suppose that m is dilation immune. From Lemma 4.4 we conclude that there exists $k \geq 1$ such that $r^{k,t} \in m^t$ for every $t \in (0, 1)$. It follows that $r^{z,t} \in m$ for every $t \in (0, 1)$ and every $z \in [1, k]$. Let $f \in \gamma_k$. By Lemma 4.7, f is a mixture of the $r^{z,t}$'s so that $f \in m$. Hence, $\gamma_k \subset m$.

Now choose $f \in m$. Suppose that $f \notin \gamma_k$. Then $\rho(f) = f^*(0)/f^*(1) = x > k$. Let $t \in (0, \frac{1}{2})$ and choose an integer n such that $(1 - 2t)/(2t) < n \leq 1/(2t)$. Define $W_1, \dots, W_n$, $Y_1, \dots, Y_n$, $B_1, \dots, B_n$ and A as in Lemma 4.3 and define h_i^t by

$$h_i^t(\omega) = \begin{cases} \dfrac{\int_0^t f^*}{t}, & \text{if } \omega \in W_i, \\ \dfrac{\int_t^{1-t} f^*}{1 - 2t}, & \text{if } \omega \in \Omega - B_i, \\ \dfrac{\int_{1-t}^1 f^*}{t}, & \text{if } \omega \in Y_i. \end{cases}$$

Then $h_i^t \in m$ since $h_i^t \prec f$. For t sufficiently small, $\rho(h_i^t) > k$. Then, by the same argument used in Lemma 4.3, $\{B_1, \dots, B_{n+1}\} \rightsquigarrow A$—a contradiction. Thus, f must be in γ_k and we conclude that $m = \gamma_k$.

Finally, we show that γ_k is dilation immune. Consider any A. Let $A = \mu(A)$. Choose a partition $\mathscr{B} = \{B_1, \dots, B_n\}$. Let $W_i = B_i \cap A$ and $Y_i = B_i \cap A^c$. It can be shown that $\overline{P}(A|B_i) = k\mu(W_i)/(k\mu(W_i) + \mu(Y_i))$ and $\underline{P}(A|B_i) = \mu(W_i)/(\mu(W_i) + k\mu(Y_i))$. Also, $\overline{P}(A) = ka/(ka + (1-a))$ and $\underline{P}(A) = a/(a + k(1-a))$. If this partition dilates A, then, for $i = 1, \dots, n$, $\overline{P}(A|B_i) \geq \overline{P}(A)$ which implies that $\mu(Y_i)/\mu(W_i) \leq (1-a)/a$. Similarly, $\underline{P}(A|B_i) \leq \underline{P}(A)$ which implies that $\mu(Y_i)/\mu(W_i) \geq (1-a)/a$. But for a dilation to occur, at least one set of these inequalities must be strict. This is impossible. Hence, there can be no dilation. □

We have shown that the only dilation immune neighborhoods of μ are density ratio neighborhoods. Recall that, for Theorem 4.1, m is assumed to be a set of density functions which rules out total variation neighborhoods and ε-contamination neighborhoods since these neighborhoods contain measures without densities. But, following Section 3, it is easy to see that these neighborhoods are dilation prone. Another class of densities, called density bounded classes, is discussed in Lavine (1991). Our theorem shows that this class is dilation prone.

5. Ramifications of dilation. In Bayesian robustness neighborhoods of probability measures are often used—see Berger (1984, 1990) and Lavine (1991). Unless density ratio classes are used, and most often they are not, the robust Bayesian must accept dilation. When sets of probabilities result from incomplete elicitation of probabilities, one response to dilation might be to elicit more precise probabilities. But the results in Section 4 show that this will not prevent dilation. Generally, it is the form of the neighborhood, not its size, that causes dilation. This does not mean that dilation prone neighborhoods should be abandoned. The fact that a dilation may occur for some event may not be a problem. However, it is important to draw attention to the phenomenon. We believe that few people who use robust Bayesian techniques are aware of the issue.

Dilation also has ramifications in decision theory. Specifically, Seidenfeld (1991) shows that dilation causes the usual relationship between normal and extensive forms of decision problems to break down.

Dilation provides an alternate axiomatic basis for precise probabilities. A precise probability may be regarded as a complete order on the set of all random variables (gambles). Many critics of probability theory insist on weakening the complete order and using, instead, a partial order. This leads to upper and lower probabilities [Walley (1991)]. Now suppose we start with a partial order and add two more axioms. The first is the existence of independent coin flips. Second, suppose we demand dilation immunity as an axiom. Then, by Theorem 2.4, the upper and lower probabilities must agree (as long as we rule out the trivial case that dilation is avoided by having upper probabilities equal to 1 and lower probabilities equal to 0). Hence, the theory of upper and lower probabilities must either tolerate dilation or must rule out all but precise probabilities. In defense of the former, see Walley (1991), page 299.

To conclude, we mention some open questions that we are currently investigating. First, we are exploring the counterparts of Lemma 3.1 and Theorem 4.1 applied to more general convex sets. For example, under what conditions will there be dilation if the convex set of probabilities is generated by a precise likelihood and a set of prior probabilities? Second, we are considering whether the "coarsening result" of Theorem 3.1 applies more generally. Third, we are investigating statistical applications of dilation. For example, Seidenfeld (1981) and Walley (1991), page 299, argue that randomization in experimental design can be understood in terms of sets of probabilities. But then dilation occurs when the ancillary data of the outcome of the randomization are known. Finally, the special role that independence plays in dilation (Section 2) suggests that there may be useful relations between dilation and the analysis of contingency tables using imprecise probabilities.

Acknowledgments. The authors thank Timothy Herron and the referees for extensive comments and suggestions.

REFERENCES

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

BERGER, J. (1985). *Statistical Decision Theory*, 2nd ed. Springer, New York.

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

BERGER, J. and BERLINER, M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

DEROBERTIS, L. and HARTIGAN, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9** 235–244.

DE FINETTI, B. (1964). Foresight: its logical laws, its subjective sources. In *Studies in Subjective Probability* (H. E. Kyburg and H. E. Smokler, eds.). Wiley, New York.

FINE, T. L. (1988). Lower probability models for uncertainty and nondeterministic processes. *J. Statist. Plann. Inference* **20** 389–411.

GOOD, I. J. (1952). Rational decisions. *J. Roy. Statist. Soc. Ser. B* **14** 107–114.

GOOD, I. J. (1966). On the principle of total evidence. *British J. Philos. Sci.* **17** 319–321.

GOOD, I. J. (1974). A little learning can be dangerous. *British J. Philos. Sci.* **25** 340–342.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd, ed. Cambridge Univ. Press.

HUBER, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Inst. Internat. Statist.* **45** 181–191.

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

HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

KOOPMAN, B. O. (1940). The bases of probability. *Bull. Amer. Math. Soc.* **46** 763–774.

KYBURG, H. (1961). *Probability and the Logic of Rational Belief*. Wesleyan Univ. Press, Middleton, CT.

KYBURG, H. (1977). Randomness and the right reference class. *J. Philos.* **74** 501–520.

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

LEVI, I. (1974). On indeterminate probabilities. *J. Philos.* **71** 391–418.

LEVI, I. (1977). Direct inference. *J. Philos.* **74** 5–29.

LEVI, I. (1982). Conflict and social agency. *J. Philos* **79** 231–247.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

RYFF, J. (1965). Orbits of L^1 functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* **117** 92–100.

SAVAGE, L. J. (1972). *The Foundations of Statistics* 2nd ed. Dover, New York.

SCHERVISH, M. J., SEIDENFELD, T. and KADANE, J. (1984). The extent of non-conglomerability of finitely additive probabilities. *Z. Wahrsch. Verw. Gebiete* **66** 205–226.

SEIDENFELD, T. (1981). Levi on the dogma of randomization. In *Henry E. Kyburg and Isaac Levi* (R. Bogdan, ed.) 263–291. Reidel, Dordrecht.

SEIDENFELD, T. (1991). When normal and extensive forms differ. *Proc. Ninth Internat. Congress of Logic, Methodology and the Philosophy of Science* (D. Pawitz, B. Skyrms and D. Westerståhl, eds.). North Holland, Amsterdam. To appear.

SEIDENFELD, T., KADANE, J., and SCHERVISH, M. (1989). On the shared preferences of two Bayesian decision makers. *J. Philos.* **86** 225–244.

SEIDENFELD, T., SCHERVISH, M. and KADANE, J. (1990). Decisions without ordering. *Acting and Reflecting* (W. Sieg. ed.) 143–170. Kluwer, Dordrecht.

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

WALLEY, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

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

WASSERMAN, L. (1992). Invariance properties of density ratio priors. *Ann. Statist.* **20** 2177–2182.

WASSERMAN, L. A. and KADANE, J. B. (1990). Bayes' theorem for Choquet capacities. *Ann. Statist.* **18** 1328–1339.

WASSERMAN, L. and KADANE, J. B. (1992). Symmetric upper probabilities. *Ann. Statist.* **20** 1720–1736.

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213