DISTENTION FOR SETS OF PROBABILITIES

RUSH T. STEWART[†] AND MICHAEL NIELSEN[‡]

† *MCMP, LMU Munich*

‡ *Department of Philosophy, The University of Sydney*

ABSTRACT. A prominent pillar of Bayesian philosophy is that, relative to just a few constraints, priors "wash out" in the limit. Bayesians often appeal to such asymptotic results as a defense against charges of excessive subjectivity. But, as Seidenfeld and coauthors observe, what happens in the short run is often of greater interest than what happens in the limit. They use this point as one motivation for investigating the counterintuitive short run phenomenon of dilation since, it is alleged, "dilation contrasts with the asymptotic merging of posterior probabilities reported by Savage (1954) and by Blackwell and Dubins (1962)" (Herron et al., 1994). A partition *dilates* an event if, relative to every cell of the partition, uncertainty concerning that event *increases*. The measure of uncertainty relevant for dilation, however, is not the same measure that is relevant in the context of results concerning whether priors wash out or "opinions merge." Here, we explicitly investigate the short run behavior of the metric relevant to merging of opinions. As with dilation, it is possible for uncertainty (as gauged by this metric) to increase relative to every cell of a partition. We call this phenomenon *distention*. It turns out that dilation and distention are orthogonal phenomena.

Keywords. Dilation; distention; imprecise probabilities; merging of opinions; polarization; total variation distance; uncertainty

1. INTRODUCTION

A specter is haunting the theory of imprecise probabilities—the specter of dilation.[1] When dilation occurs, learning new information *increases* uncertainty. Dilation is especially interesting because, relative to a dilating partition, uncertainty grows *no matter* which cell an agent learns. This has prompted investigations into the rational status of willingness to pay "negative tuition," that is, willingness to pay not to learn (e.g., Kadane et al., 2008). Yet dilation is not the only way for uncertainty to grow relative to every cell of a partition for imprecise probabilities (IP). With dilation, the focus is on the uncertainty about a particular event. But uncertainty about a given event is not the only kind of uncertainty with which we might be concerned. We might instead be concerned about *overall* uncertainty. In this study, we will be so concerned. Given a set of probabilities and a (positive, measurable) partition, *distention* occurs when the (supremum of the) *total variation distance* increases *no matter* which cell of the partition an agent learns. Since each cell induces an increase in total variation for a set of probabilities, conditional on any cell, the set of probabilities is "more

[1]We are delighted to contribute to this collection of essays in honor of Teddy Seidenfeld's career. We have both learned a tremendous amount from Teddy's work, which is a model of clear and precise philosophy. We hope the present paper contributes to one of the many interesting topics of Teddy's research.

spread" than it is unconditionally. In this sense, uncertainty—not about a particular event, but of a global sort—is sure to grow. Distention, like dilation, then, is a way for evidence to increase uncertainty across an entire evidential partition. As far as we know, ours is the first articulation and investigation of the phenomenon of distention.

Several considerations motivate our study. With their justly celebrated "merging of opinions" theorem, Blackwell and Dubins establish that, relative to just a few assumptions, Bayesians achieve consensus in the limit almost surely (1962). That priors "wash out" in this way is an important pillar of Bayesian philosophy (Savage, 1954; Edwards et al., 1963; Gaifman and Snir, 1982; Earman, 1992; Huttegger, 2015).[2] Schervish and Seidenfeld extend Blackwell and Dubins's result to IP theory, establishing that certain convex polytopes of probabilities exhibit uniform merging (Schervish and Seidenfeld, 1990, Corollary 1).[3] But as Herron, Seidenfeld, and Wasserman observe about Blackwell and Dubins's result, "What happens asymptotically, almost surely, is not always a useful guide to the short run" (1997, p. 412). Disagreements can persist, or even increase, over finite time horizons even though they vanish in the limit. Herron et al. use this point, however, to motivate an investigation into dilation. The idea is supposed to be that an increase in disagreement among the elements of a set of probabilities in the dilation sense is the opposite of an increase in agreement among those elements in the merging sense. But, as we will show, an occurrence of dilation does not imply an increase in disagreement in the Blackwell and Dubins model (Section 4). We propose instead to investigate the "short run" behavior of total variation, the metric with which Blackwell and Dubins are concerned. One way of reading our position in this paper is that much of the attention bestowed on dilation amounts to stolen valor.

Another motivation for investigating distention comes from social epistemology. In Nielsen and Stewart (2020), we introduce the notions of local and global probabilistic opinion polarization between agents. There, we note 1) that the dilation phenomenon for imprecise probabilities is in some ways analogous to local polarization, and 2) that local and global polarization are logically independent. This presents our context of discovery for distention: it is the phenomenon analogous to global polarization for imprecise probabilities.

Furthermore, in many cases, it is natural to be concerned with overall uncertainty as we construe it in this essay. Many inquiries do not center on just a single event or proposition of interest, but focus on a host of questions. At least, we claim, this is one legitimate way to construe some inquiries. For such inquiries, an agent or group may be concerned with his or their estimates over an entire space of possibilities, and with how new information affects those estimates. In this kind of case, total variation seems the more appropriate measure of increases and decreases of uncertainty.

After rehearsing the basics of dilation (Section 2), we define distention precisely (Section 3), show that it is logically independent of dilation (Section 4, Proposition 1), and provide a characterization (Section 5, Proposition 2). We then draw some connections between local and global polarization in social epistemology, on the one hand, and dilation and distention in IP theory, on the other (Section 6). We conclude by considering some further ramifications of distention (Section 7).

---

[2]As Edwards, Lindman, and Savage write, "This approximate merging of initially divergent opinions is, we think, one reason why empirical research is called 'objective'" (1963, p. 197).

[3]Convexity is often imposed on sets of probabilities in the IP setting (e.g., Levi, 1980). Convex polytopes of probabilities emerge naturally in many contexts for IP (e.g., Levi, 1985; Stewart and Ojea Quintana, 2018), with Bronevich and Klir even claiming that "It is convenient and rational [...] that each such set of probability measures is a convex polytope" (2010, p. 366). We return briefly to the topic of convexity in Section 5.

2. DILATION

Our main interest in this essay is in certain aspects of the theory of imprecise probabilities. We adopt a formalism based on *sets of probability measures*, though several alternative frameworks have been studied (Walley, 2000; Augustin et al., 2014). There are a number of motivations for IP. Imprecise probabilities are an important tool in robustness analysis for standard Bayesian inference (Walley, 1991; Berger, 1994). Sets of probabilities are useful in studying group decision problems (Levi, 1982; Seidenfeld et al., 1989) and opinion pooling (Elkin and Wheeler, 2018; Stewart and Ojea Quintana, 2018). IP provides more general models of uncertainty which are often championed as superior for a number of normative considerations relevant to epistemology and decision making (Levi, 1974; Walley, 1991). Sets of probabilities can also be used to represent partial elicitation of precise subjective probabilities. Some have argued that IP presents a more *realistic* theory of human epistemology (Arló-Costa and Helzner, 2010). IP allows for a principled introduction of incomplete preferences in the setting of expected utility maximization (Seidenfeld, 1993; Kaplan, 1996), and has been used to offer resolutions of some of the paradoxes of decision (Levi, 1986). And there are other considerations driving the development of the theory of imprecise probabilities.

Dilation is the (at least at first blush) counterintuitive phenomenon of learning *increasing* uncertainty.[4] For a dilating partition, learning *any* cell results in greater uncertainty. Take the simple, stock example of flipping a coin. This experiment partitions the sample space into two cells, one corresponding to heads, the other to tails. It could be the case that, for some event A, *no matter how the coin lands*, the agent's estimate for A ($P(A) = 0.5$, say) will be strictly included in the agent's estimate conditional on the outcome of the coin toss ([0.1, 0.9], for example). Example 1 details such a case.

Throughout, let Ω be a sample space of elementary events or possible worlds. Elements of Ω can be thought of as maximally specific epistemic possibilities for an agent. Let $\mathcal{F}$ be a sigma-algebra on Ω, i.e., a non-empty collection of subsets of Ω closed under complementation and countable unions. Elements of $\mathcal{F}$ are called *events* and $\mathcal{F}$ can be thought of as a general space of possibilities (not just maximally specific ones). We assume the standard ratio definition of conditional probability:

$$P(A|E) = \frac{P(A \cap E)}{P(E)}, \text{ when } P(E) > 0.$$

Let $\mathbb{P}$ be a *set* of probability measures. Such a set can be interpreted, for example, as the probability measures an agent regards as permissible to use in inference and decision problems, those distributions he hasn't ruled out for such purposes. If $\mathbb{P}$ is convex, it associates with any event in the algebra an interval of probability values (such as $[0.1, 0.9]$).[5] We can now define dilation precisely.

Definition 1. Let $\mathbb{P}$ be a set of probabilities on $(\Omega, \mathcal{F})$, let $\mathscr{B}$ be a positive partition of Ω[6], and let $A \in \mathcal{F}$. We say that the partition $\mathscr{B}$ *dilates* A just in case, for each $E \in \mathscr{B}$,

$$\inf\{P(A|E) : P \in \mathbb{P}\} < \inf\{P(A) : P \in \mathbb{P}\} \leq \sup\{P(A) : P \in \mathbb{P}\} < \sup\{P(A|E) : P \in \mathbb{P}\}.$$

---

[4]There is by now a fairly extensive literature on dilation (e.g., Walley, 1991; Seidenfeld and Wasserman, 1993; Herron et al., 1994; Wasserman and Seidenfeld, 1994; Herron et al., 1997; Pedersen and Wheeler, 2014, 2015; Nielsen and Stewart, 2019).

[5]We call $\mathbb{P}$ *convex* when $P, Q \in \mathbb{P}$ implies $aP + (1-a)Q \in \mathbb{P}$ for every $a \in [0,1]$. The *convex hull* of a set of points is the smallest convex set containing those points.

[6]The partition $\mathscr{B}$ is positive if $E \in \mathscr{B}$ implies $E \in \mathcal{F}$ and $P(E) > 0$ for all $P \in \mathbb{P}$. Note that this definition entails that every cell of $\mathscr{B}$ is measurable. Also note that positive partitions are necessarily countable.

It is clear that precise credal states are dilation-immune since $\inf\{P(H|E)\} = \sup\{P(H|E)\}$ for all H and E in $\mathcal{F}$ such that $P(H|E)$ is defined.

Consider the following common example of dilation, introduced in outline earlier (Herron et al., 1994; Pedersen and Wheeler, 2015). We simplify by assuming that $\mathbb{P}$ consists of just two probabilities.

Example 1. *Let $\mathbb{P} = \{P_1, P_2\}$ be a set of probabilities on $(\Omega, \mathcal{F})$. Suppose that, for $G \in \mathcal{F}$, $P_1(G) = 0.1$ and $P_2(G) = 0.9$. Relative to $\mathbb{P}$, then, G is a highly uncertain event. Consider the toss of a coin that is fair according to both P_1 and P_2: $P_1(H) = P_2(H) = 1/2 = P_1(H^c) = P_2(H^c)$. Suppose that the outcomes of the coin toss are independent of the event G according to both P_1 and P_2. Then, $P_1(G \cap H) = P_1(G)P_1(H)$ and $P_2(G \cap H) = P_2(G)P_2(H)$. Let A be the "matching" event that either both G and H occur or both do not. That is, $A := (G \cap H) \cup (G^c \cap H^c)$. Notice that $P_1(A) = 1/2 = P_2(A)$. Despite initial agreement concerning A, the coin toss dilates P_1 and P_2 on A. For $i \in \{1, 2\}$,*

$$P_i(A|H) = \frac{P_i([(G \cap H) \cup (G^c \cap H^c)] \cap H)}{P_i(H)} = \frac{P_i(G \cap H)}{P_i(H)} = \frac{P_i(G)P_i(H)}{P_i(H)} = P_i(G).$$

So even though both P_1 and P_2 assign probability $1/2$ to A initially, learning that the coin lands heads yields $P_1(A|H) = 0.1$ and $P_2(A|H) = 0.9$. Hence, $P_1(A|H) < P_1(A) \le P_2(A) < P_2(A|H)$. Analogous reasoning establishes that $P_2(A|H^c) < P_2(A) \le P_1(A) < P_1(A|H^c)$. △

Some see in dilation grounds for rejecting the notion that imprecise probabilities provide a normatively permissible generalization of standard Bayesian probability theory (e.g., White, 2010; Topey, 2012). It is not just that it seems intuitively wrong that learning should increase uncertainty. Dilation has further consequences. For example, dilation leads to violations of Good's Principle. Good's Principle enjoins us to delay making a terminal decision if presented with the opportunity to first learn cost-free information. For the standard, Bayesian expected utility framework, Good's Principle is backed up by a theorem. Good famously shows that, in the context of expected utility maximization, the value of making a decision *after* learning cost-free information is always greater than or equal to the value of making a decision *before* learning (Good, 1967).[7] Dilation, however, leads to the devaluation of information (e.g., Pedersen and Wheeler, 2015). With dilation, an agent may be willing to actually *pay* to forgo learning some information, what Kadane et al. label "negative tuition" (Kadane et al., 2008).

3. Distention

What would it mean for uncertainty to grow with respect to every cell of an experimental partition, though not uncertainty about a single, fixed event? We adopt the same metric that Blackwell and Dubins employ to gauge consensus in the context of merging of opinions. For any two probabilities, P_1 and P_2, the *total variation distance d* is given by

$$d(P_1, P_2) = \sup_{A \in \mathcal{F}} |P_1(A) - P_2(A)|.$$

---

[7]More precisely, the value of deciding before learning is given by the maximum expected utility of the options. That value is always less than or equal to the expected value of the maximum expected utility of the options after learning, where expected utility after learning is calculated with the relevant conditional probability.

When $d(P_1, P_2) = 0$, it follows that $P_1 = P_2$. And if P_1 and P_2 are within ϵ according to d, they are within ϵ for every event in the algebra. We will have occasion to appeal to the fact that, in finite probability spaces, the total variation distance is given by

$$d(P_1, P_2) = P_1(A_0) - P_2(A_0), \tag{1}$$

where $A_0 = \{\omega \in \Omega : P_1(\omega) > P_2(\omega)\}$ (e.g., Nielsen and Stewart, 2020).

So we take it that for global uncertainty to grow with respect to each cell of an experimental partition is for the total variation to increase conditional on each cell.[8] That, in turn, means that, for every cell, there is some event such that the "distance" between the probabilities for that event conditional on that cell is greater than the distance between probabilities for any event unconditionally. For an arbitrary set of probabilities, we look at the supremum of the total variation for all elements of the set. To simplify notation, let us adopt some metric space terminology and call $d(\mathbb{P}) = \sup_{P,Q \in \mathbb{P}} d(P, Q)$ the *diameter* of $\mathbb{P}$. If $P(E) > 0$ for all $P \in \mathbb{P}$, then let us write $\mathbb{P}^E = \{P^E : P \in \mathbb{P}\}$, where $P^E = P(\cdot \mid E)$. We should stress that whenever we write $\mathbb{P}^E$, we are assuming that all $P \in \mathbb{P}$ assign E positive probability.

Definition 2. Let $\mathbb{P}$ be a set of probabilities on $(\Omega, \mathcal{F})$, let $\mathscr{B}$ be a positive partition of Ω. We say that the partition $\mathscr{B}$ *distends* $\mathbb{P}$ just in case, for each $E \in \mathscr{B}$,

$$d(\mathbb{P}) < d\left(\mathbb{P}^E\right).$$

Another way to think of distention is that a partition that distends $\mathbb{P}$ pushes the elements of $\mathbb{P}$ further from consensus. When $\mathbb{P}$ is interpreted as the credal state of a single agent, the closer a set of probabilities gets to "consensus," the closer it is that uncertainty is reduced to risk— a unique probability function—for an agent. So distention pushes uncertainty further from being reduced to simple risk. Like dilation, then, distention is a way that uncertainty grows whatever the outcome of an experiment. Unlike dilation, though, the focus for distention is on total variation distance and not the probability of a single, fixed event.

As repeatedly noted in the literature (e.g., Seidenfeld and Wasserman, 1993; Pedersen and Wheeler, 2015), dilation bears certain similarities to non-conglomerability. Let $\mathscr{B} = \{E_i : i \in I\}$ be a positive partition. We say that A is *conglomerable* in $\mathscr{B}$ when

$$\inf\{P(A|E) : E \in \mathscr{B}\} \leq P(A) \leq \sup\{P(A|E) : E \in \mathscr{B}\}.$$

And we say that P is conglomerable in $\mathscr{B}$ if the above inequalities hold for all events A. When A is non-conglomerable in $\mathscr{B}$, $P(A)$ cannot be regarded as a weighted average of the probabilities $P(A|E_i)$. If $\mathscr{B}$ is a countable partition and P is not conglomerable for A in $\mathscr{B}$, then the law of total probability fails. This happens only when P fails to be countably additive. Schervish et al. prove that, for any merely finitely additive probability P (on a space admitting a countably infinite partition), there is some event A and countable partition $\mathscr{B}$ such that P fails to be conglomerable for A in $\mathscr{B}$ (1984). One reason non-conglomerability is odd is because it allows for reasoning to foregone conclusions (Kadane et al., 1996). Merely running an experiment, regardless of the outcome, allows one to uniformly increase (or decrease) one's estimate in some event. In other words, an experiment could be designed such that, before even running it, the experimenter can be sure that conditionalizing on the outcome will yield a higher (or lower, depending on the case) probability for the event in question. Like dilation, non-conglomerability also leads to the devaluation of information in violation of Good's Principle (e.g., Pedersen and Wheeler, 2015). Distention, like dilation but unlike

---

[8]See subsection 7.3 for some comments on alternative interpretations of increasing global uncertainty.

non-conglomerability, can occur even on finite sets. So, like dilation, but perhaps unlike non-conglomerability, distention cannot be explained away by poor intuitions concerning infinite sets.

4. Distention Is Logically Independent of Dilation

Given certain conceptual similarities between distention and dilation, it is natural to ask about their logical relations. The answer to that query is that dilation does not imply distention, nor does distention imply dilation. In other words, dilation and distention are logically independent.

To see that dilation does not imply distention, return to the coin example from earlier.

Example 2. *Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $A = \{\omega_1, \omega_2\}$, and $H = \{\omega_1, \omega_4\}$. Let $\mathbb{P} = \{P, Q\}$, given on the following table along with their updates on H and on H^c.*

Table 1. Dilation without Distention

	ω_1	ω_2	ω_3	ω_4
P	0.05	0.45	0.05	0.45
Q	0.45	0.05	0.45	0.05
P^H	0.1	0	0	0.9
Q^H	0.9	0	0	0.1
P^{H^c}	0	0.9	0.1	0
Q^{H^c}	0	0.1	0.9	0

Take $\mathscr{B} = \{H, H^c\}$ as our experimental partition (the outcome of a flip of a fair coin). From the table we compute $P(A) = 0.5 = Q(A)$. Yet, $P(A|H) = 0.1$ and $Q(A|H) = 0.9$. Similarly, $P(A|H^c) = 0.9$ and $Q(A|H^c) = 0.1$. So, $\mathscr{B}$ dilates $\mathbb{P}$ on A. However, again computing from the table using Equation 1, we have $d(P, Q) = d\left(P^H, Q^H\right) = d\left(P^{H^c}, Q^{H^c}\right) = 0.8$. It follows that dilation does not entail distention. $\triangle$

To see that distention does not imply dilation, consider the following simple example.

Example 3. *Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $H = \{\omega_1, \omega_2\}$, and $\mathbb{P} = \{P, Q\}$, given on Table 2. Consider the partition $\mathscr{B}$ consisting of H and its complement.*

Table 2. Distention without Dilation

	ω_1	ω_2	ω_3	ω_4
P	1/10	1/5	1/10	3/5
Q	1/10	1/10	1/5	3/5
P^H	1/3	2/3	0	0
Q^H	1/2	1/2	0	0
P^{H^c}	0	0	1/7	6/7
Q^{H^c}	0	0	1/4	3/4

While $d(P, Q) = 1/10$, $d\left(P^H, Q^H\right) = 1/6$, and $d\left(P^{H^c}, Q^{H^c}\right) = 3/28$. So $\mathscr{B}$ distends $\mathbb{P}$. But it does not dilate any event. Not only is there no dilation in $\mathscr{B}$, no partition of Ω dilates any event. This can be checked, a bit tediously, by hand.[9] $\triangle$

A set of probabilities cannot exhibit distention on a smaller sample space. That is because any (non-trivial) partition on a smaller space will have a singleton as a cell. In that case, provided the partition is positive, the distance between probabilities conditional on a singleton is 0.

We submit that the short run that is relevant to merging of opinions is the short run behavior of total variation distance and not the sort of behavior exemplified by dilation. After all, it is the total variation distance that Blackwell and Dubins use to measure consensus. Examples 2 and 3 show that dilation is in fact orthogonal to distention, but Example 4 shows that distention and dilation in a given partition are consistent (see the Appendix). We summarize these findings in the following proposition.

Proposition 1. *While a set $\mathbb{P}$ can exhibit both dilation and distention simultaneously with respect to a single partition, dilation does not imply distention, nor does distention imply dilation.*

5. A CHARACTERIZATION OF DISTENTION

For any two probabilities P and Q and any two events A and E such that $P(E), Q(E) > 0$, define a function $\bar{\mathcal{B}}$ as follows:

$$\bar{\mathcal{B}}_{P,Q}(A, E) = \frac{P(A)}{P(E)} - \frac{Q(A)}{Q(E)}. \tag{2}$$

In a way, the function $\bar{\mathcal{B}}$ sets the so-called *Bayes factor* in difference form. The Bayes factor for P and Q with respect to A and E is defined as

$$\mathcal{B}_{P,Q}(A, E) = \frac{P(A)}{P(E)} \bigg/ \frac{Q(A)}{Q(E)}. \tag{3}$$

Bayes factors have a distinguished pedigree in Bayesian thought (Good, 1983; Wagner, 2002; Jeffrey, 2004). Wagner, for instance, contends that identical learning experiences for two agents are captured by identical Bayes factors for their respective priors and posteriors rather than by identical posterior opinions. But $\bar{\mathcal{B}}$ differs substantially in interpretation from a Bayes factor. In particular, it is not assumed that either of P or Q is an update of the other.

The function $\bar{\mathcal{B}}$ allows us to state one simple characterization of distention. Since convexity has played a prominent role in IP theory, we also state an equivalence with the distention of the convex hull.[10]

Proposition 2. *Let $\mathbb{P}$ be a set of probabilities on $(\Omega, \mathcal{F})$, and let $\mathscr{B}$ be a positive partition of Ω. The following are equivalent.*

  (I) *$\mathscr{B}$ distends $\mathbb{P}$.*
 (II) *For all $E \in \mathscr{B}$ there exist $P, Q \in \mathbb{P}$, and $A \subseteq E$ such that*

$$\bar{\mathcal{B}}_{P,Q}(A, E) > d(\mathbb{P}). \tag{4}$$

(III) *$\mathscr{B}$ distends the convex hull of $\mathbb{P}$.*

---

[9]Note, though, that only partitions consisting of non-singleton cells, of which there are just three, need to be checked. Dilation will be thwarted by any partition containing a singleton because the resulting conditional distributions will agree.

[10]For one important debate about the normative status of convexity for IP, see (Levi, 1980; Seidenfeld et al., 1989; Levi, 1990; Seidenfeld et al., 2010; Levi, 2009).

We regard Proposition 2 as a first pass at characterizing distention. The problem of finding such characterizations is more than a purely formal one. The characterizing conditions should be relatively simple and provide insight into the "wherefore" of distention. It is not clear to us that Proposition 2 satisfies the second desideratum.

6. Local and Global Polarization

Polarization is a social phenomenon. Accordingly, in our previous related study (2020), we were concerned about its implications for social epistemology. But, as we noted there, social epistemology and the theory of imprecise probability gain much from cross-fertilization. In this paper, we exploit concepts from social epistemology in the hopes of gaining a deeper understanding of the theory of imprecise probabilities.

Like dilation, *local* polarization is defined in terms of a specific event. Polarization in this sense occurs when shared evidence pushes opinions about a specific event further apart.

Definition 3. Let P_1 and P_2 be probability functions on $(\Omega, \mathcal{F})$, and let $A, E \in \mathcal{F}$. We say that evidence E *polarizes* P_1 *and* P_2 *with respect to the event* A if

$$P_1(A|E) < P_1(A) \le P_2(A) < P_2(A|E).$$

The possibility of two agents polarizing when updating on shared evidence may itself come as a surprise to some. In particular, the fact that it is possible for Bayesians to polarize is a challenge to the view that rational agents who share evidence resolve disagreements. Elsewhere, we have labeled this view *The Optimistic Thesis about Learning* (TOTAL) and, at Gordon Belot's suggestion, its proponents TOTALitarians (2020). Such a view seems to underwrite many of our ordinary practices (in rational persuasion, advocacy, etc.) as well as positions in current philosophical debates. For example, the view that an epistemic peer's disagreement is evidence of defect in one's own beliefs, as some so-called conciliationists allege, seems committed to TOTAL. Bayesian polarization, however, suggests TOTAL is false.

Not only does the definition of local polarization resemble that of dilation, local polarization and dilation can be characterized in terms of similar conditions (cf. Seidenfeld and Wasserman, 1993, Result 1; Nielsen and Stewart, 2020, Theorem 1). But we can be more precise than mere resemblance. Let $\mathbb{P} = \{P_1, P_2\}$ and let $\mathcal{B}$ be a positive *finite* partition that dilates A. Then there is some $E \in \mathcal{B}$ such that E polarizes P_1 and P_2 with respect to A. If not, then dilation implies that

$$P_1(A) \le P_2(A) < P_1(A \mid E)$$

for all $E \in \mathcal{B}$, where we have assumed the first inequality without loss of generality. Multiplying by $P_1(E)$ and summing over $E \in \mathcal{B}$ yields

$$P_1(A) = \sum_{E \in \mathcal{B}} P_1(A)P_1(E) < \sum_{E \in \mathcal{B}} P_1(A \mid E)P_1(E) = P_1(A),$$

which is a contradiction. Hence, dilation guarantees that some cell of the dilating partition is polarizing.

Central to the concept of *global* polarization is a measure of the extent of total disagreement between two probability functions. Again, we adopt the total variation metric to assess total disagreement. Naturally enough, we say that global polarization occurs when shared evidence brings about an increase in total variation between two probability functions.

Definition 4. Evidence E *polarizes* P_1 *and* P_2 *globally* if $d(P_1, P_2) < d\left(P_1^E, P_2^E\right)$.

In contrast to the optimistic spin typically put on the Blackwell-Dubins merging result, our *consensus-or-polarization law* shows that even very mild and plausible weakenings of the relevant assumptions no longer entail almost sure consensus in the limit. Rather, agents achieve consensus or maximally (globally) polarize with probability 1 (Nielsen and Stewart, 2020, Theorem 3).

Local and global polarization are logically independent. While probabilities can exhibit local and global polarization simultaneously, global polarization does not imply local polarization, nor does local polarization imply global polarization (Nielsen and Stewart, 2020, Proposition 1). As we saw above, the IP analogues of local and global polarization, dilation and distention, respectively, exhibit the same sort of logical independence.

7. Some Upshots

7.1. **Asymptotic Consensus.** The primary precondition of Blackwell and Dubins' merging theorem is *absolute continuity*.[11] If P is absolutely continuous with respect to Q, then $Q(A) = 0$ implies $P(A) = 0$ for all $A \in \mathcal{F}$. Their theorem roughly says that if P is absolutely continuous with respect to Q, then P assigns probability 1 to achieving consensus with Q in the limit. The examples above involve regular prior distributions on finite probability spaces. Every probability function is absolutely continuous with respect to a regular distribution. In larger spaces, regularity is not achievable. This makes the issue of absolute continuity nontrivial. Extending the theorem to sets of probability functions presents further complications. Schervish and Seidenfeld establish that closed, convex sets of mutually absolutely continuous probabilities that are generated by finitely many extreme points merge under Bayesian conditionalization (Schervish and Seidenfeld, 1990, Corollary 1). In previous work, we generalize this result, showing that closed, convex sets of mutually absolutely continuous probabilities that are generated by finitely many extreme points merge under Jeffrey conditioning as well (Stewart and Nielsen, 2019, Proposition 1).[12] For such sets of distributions, the significance of distention depends on the importance of the short run. In our opinion, the importance is clear. For all Blackwell and Dubins's theorem says, approximate consensus may be achieved only in the *very* long run. Many things for which consensus is relevant happen in the not very long run. Even if $\mathbb{P}$ is a set of mutually absolutely continuous probabilities (and so subject to the merging theorem), not only may its elements fail to achieve consensus in the short run, they might collectively distend, moving *away* from consensus whatever evidence comes. Of course, if an IP set does not consist of mutually absolutely continuous priors, failure of almost sure asymptotic consensus is a foregone conclusion.

7.2. **Group Manipulation.** Moving now to the social setting, distention implies the possibility of a sort of group manipulation in the short run. Interpret a set $\mathbb{P}$ as the (individually precise) probabilities of a group of agents. For certain such sets, an experiment can be designed such that, no matter the outcome, the group will be *further* from consensus as a result of learning *shared* evidence. If a policy decision or group choice requires consensus (or a

---

[11]The theorem also assumes that probabilities admit regular conditional distributions (Billingsley, 2008). A stronger assumption that implies the existence of regular conditional distributions is that all sub-sigma-algebras of the filtration are generated by countable partitions. This assumption is used, for example, in (Kalai and Lehrer, 1994).

[12]One interesting thing about the generalization to Jeffrey conditioning is that, unlike standard Bayesian conditionalization, Jeffrey conditioning does not generally preserve the convexity of the initial set (Stewart and Nielsen, 2019, Proposition 3). Another is that "uncertain learning" has been rarely married with general models of uncertainty along IP lines in the literature.

tolerance of only ϵ disagreement) on some algebra of events, such decision making can be frustrated (at least in the short run) by a devious experimenter no matter the outcome of the experiment.

7.3. **Alternative Measures of Uncertainty.** We have focused on total variation distance because of its distinguished role in merging of opinions and, consequently, Bayesian thought, and because of merging's alleged contrast with dilation. Total variation, however, is one example of a large class of *divergences* between probabilities known as *f-divergences*. Another prominent example is *Kullback-Leibler* (KL) divergence from Q to P defined in discrete spaces by

$$D_{KL}(P||Q) = -\sum_i P(\omega_i) \log \frac{Q(\omega_i)}{P(\omega_i)}.$$

An important fact about KL divergence, often pointed out, is that, unlike total variation, KL divergence is not a true metric. For instance, it is not symmetric. Above, we provided an example of distention without dilation (Example 3). This example also establishes that distention does not imply that the KL divergence increases across the partition. In particular, $D_{KL}(P||Q) > D_{KL}(P^E||Q^E)$, as can easily be computed from Table 2. Still other, IP-specific measures of uncertainty have been explored in the literature (e.g., Bronevich and Klir, 2010). Absent strong reasons to privilege some such measure over the others—and perhaps there are such reasons for total variation—these simple observations urge caution in drawing general lessons from dilation- or distention-type phenomena.

7.4. **"Pathologies" of Imprecision.** The further ramifications of distention remain to be explored. As we point out above, in the social setting, distention implies the possibility of certain sorts of group manipulation. For an individual with an imprecise credal state, an analogous sort of manipulation is possible in contexts in which a precise estimate is desired. For certain credal states, an experimenter can guarantee that the agent gets further (as measured by the total variation metric) from a precise estimate no matter what. How dramatic are the consequences of this sort of manipulation? And what other sorts of surprising effects, like the violations of Good's Principle for dilation, might distention bring in tow? We hope to explore these issues in future research.

One interesting point, we find, is that none of the alleged pathologies discussed in connection with imprecise probabilities seem to be at all unique to a specific IP phenomenon, nor even unique to IP given social interpretations of sets of probabilities. Violations of Good's Principle do not require dilation. Non-conglomerability leads to such violations, too. Neither does the strange phenomenon of learning increasing uncertainty imply dilation. With distention, uncertainty increases whatever evidence comes in as well. In a social setting, dilation and distention are somewhat robbed of their apparent counterintuitive sting. The lesson there is that updating on shared evidence does not guard against various types of group opinion polarization (what could be called "social uncertainty"), as mundane examples illustrate (Nielsen and Stewart, 2020).

One might take these anomalies as an argument for restricting to precise probabilities on finite spaces—by our lights, far beyond the pale of what is warranted. For one thing, continuous random variables are essential in many scientific applications and are unavailable in finite spaces. For another, violations of Good's Principle do not require imprecise probabilities, so the restriction to precise probabilities fails as a safeguard. True, there are no instances of non-conglomerability in finite spaces, but suppose with us that the restriction to such spaces is too costly. By requiring countable additivity, one guarantees conglomerability in countable

partitions. But, depending on the theory of conditional probabilities that we adopt, even countably additive probabilities can exhibit non-conglomerability in uncountable partitions. And the moral is more general still (Schervish et al., 2017). So such proposed restrictions are costly, hasty, and ineffective.

Appendix

Example 4

The following example shows that a set $\mathbb{P}$ can exhibit both dilation and distention simultaneously with respect to a single partition.

Example 4. *Let* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $E = \{\omega_1, \omega_2\}$, *and* $A = \{\omega_1, \omega_3\}$. *We take* $\mathscr{B} = \{E, E^c\}$ *as our experimental partition.*

TABLE 3. Distention is Consistent with Dilation

	ω_1	ω_2	ω_3	ω_4
P	1/100	37/100	30/100	32/100
Q	20/100	41/100	1/100	38/100
P^E	1/38	37/38	0	0
Q^E	20/61	41/61	0	0
P^{E^c}	0	0	30/62	32/62
Q^{E^c}	0	0	1/39	38/39

Calculating the total variation distance from the table, we have $d(\mathbb{P}) = d(P, Q) = 0.29$, $d\left(P^E, Q^E\right) \approx 0.302$, *and* $d\left(P^{E^c}, Q^{E^c}\right) \approx 0.46$. *So* $\mathscr{B}$ *distends* $\mathbb{P}$. *For dilation, notice that* $Q(A) = 21/100$ *and* $P(A) = 31/100$. *But* $P^E(A) = 1/38 < 21/100 < 31/100 < 20/61 = Q^E(A)$. *Similarly,* $Q^{E^c}(A) = 1/39 < 21/100 < 31/100 < 30/62 = P^{E^c}(A)$. *So,* $\mathscr{B}$ *dilates* A.

Proof of Proposition 2

Proof. We start by showing that (I) and (II) are equivalent. Suppose that (II) holds and let $E \in \mathscr{B}$. Then there exist $P, Q \in \mathbb{P}$, and $A \subseteq E$ such that

$$d(\mathbb{P}) < \left|P^E(A) - Q^E(A)\right| \leq d\left(P^E, Q^E\right) \leq d\left(\mathbb{P}^E\right).$$

Hence, $\mathscr{B}$ distends $\mathbb{P}$, so (II) implies (I).

Conversely, suppose that $\mathscr{B}$ distends $\mathbb{P}$ and let $E \in \mathscr{B}$. Then there are $P, Q \in \mathbb{P}$ such that

$$d(\mathbb{P}) < d\left(P^E, Q^E\right).$$

Let p and q be densities for P and Q, respectively, with respect to any common dominating measure m, that is, both P and Q are absolutely continuous with respect to m. (Let $m = P/2 + Q/2$, for instance.) Define

$$p^E = \frac{\mathbf{1}_E p}{P(E)} \quad \text{and} \quad q^E = \frac{\mathbf{1}_E q}{Q(E)},$$

so that p^E and q^E are densities for P^E and Q^E with respect to m. Note that the set $A = \{\omega \in \Omega : p^E(\omega) > q^E(\omega)\}$ is a subset of E because if $\omega \notin E$, then $p^E(\omega) = 0 = q^E(\omega)$. We now have,

$$d(\mathbb{P}) < d\left(P^E, Q^E\right) = P^E(A) - Q^E(A) = \frac{P(A)}{P(E)} - \frac{Q(A)}{Q(E)} = \bar{\mathcal{B}}_{P,Q}(A, E),$$

where the first equality is the general version of (1). This establishes (II), and shows that (I) and (II) are equivalent.

Next, we show that (I) is equivalent to (III). We use the following lemmas and include proofs for the reader's convenience.

Lemma 1. *For any set of probabilities $\mathbb{P}$, $d(\mathbb{P}) = d(\mathrm{co}(\mathbb{P}))$.*

Proof of Lemma 1. Since $\mathbb{P} \subseteq \mathrm{co}(\mathbb{P})$, $d(\mathbb{P}) \leq d(\mathrm{co}(\mathbb{P}))$. To show the reverse inequality, let $P, Q \in \mathrm{co}(\mathbb{P})$ be arbitrary. Then, $P = \sum_{i=1}^{n} a_i P_i$, $Q = \sum_{j=1}^{m} b_j P_j$ for some $n, m \in \mathbb{N}$, $P_i, P_j \in \mathbb{P}$, and $a_i, b_j \geq 0$ with $\sum_i a_i = 1 = \sum_j b_j$. For all $A \in \mathcal{F}$,

$$|P(A) - Q(A)| = \left| \sum_{i=1}^{n} a_i P_i(A) - \sum_{j=1}^{m} b_j P_j(A) \right| = \left| \sum_{i=1}^{n} a_i \sum_{j=1}^{m} b_j P_i(A) - \sum_{j=1}^{m} b_j \sum_{i=1}^{n} a_i P_j(A) \right|$$

$$\leq \sum_{i=1}^{n} a_i \sum_{j=1}^{m} b_j |P_i(A) - P_j(A)| \leq \sum_{i=1}^{n} a_i \sum_{j=1}^{m} b_j d(\mathbb{P}) = d(\mathbb{P}).$$

Since this holds for all $A \in \mathcal{F}$, we have

$$d(P, Q) \leq d(\mathbb{P}).$$

And since this holds for all $P, Q \in \mathrm{co}(\mathbb{P})$, we have

$$d(\mathrm{co}(\mathbb{P})) \leq d(\mathbb{P}),$$

which proves the lemma. $\qquad\square$

Lemma 2. *For any set of probabilities $\mathbb{P}$, $\mathrm{co}\left(\mathbb{P}^E\right) = \mathrm{co}(\mathbb{P})^E$.*

Proof of Lemma 2. First, let $P \in \mathrm{co}\left(\mathbb{P}^E\right)$. Then $P = \sum_{i=1}^{n} a_i P_i(\cdot \mid E)$ for some $n \in \mathbb{N}$, $P_i \in \mathbb{P}$ and $a_i \geq 0$ with $\sum_i a_i = 1$. Let $b_i = \frac{a_i}{P_i(E)N} \geq 0$, where $N = \sum_i \frac{a_i}{P_i(E)}$ is a normalizing constant that ensures $\sum_i b_i = 1$. Then,

$$P = \sum_{i=1}^{n} a_i P_i(\cdot \mid E) = \frac{\sum_i \frac{a_i P_i(\cdot \cap E)}{N P_i(E)}}{\sum_i \frac{a_i P_i(E)}{N P_i(E)}} = \frac{\sum_i b_i P_i(\cdot \cap E)}{\sum_i b_i P_i(E)} = \left(\sum_i b_i P_i \right)(\cdot \mid E) \in \mathrm{co}(\mathbb{P})^E.$$

Hence, $\mathrm{co}\left(\mathbb{P}^E\right) \subseteq \mathrm{co}(\mathbb{P})^E$.

Next, suppose that $P \in \mathrm{co}(\mathbb{P})^E$. Then $P = \left(\sum_{i=1}^{n} a_i P_i \right)(\cdot \mid E)$ for some $n \in \mathbb{N}$, $P_i \in \mathbb{P}$ and $a_i \geq 0$ with $\sum_i a_i = 1$. Let $b_i = \frac{a_i P_i(E)}{N}$ and $N = \sum_i a_i P_i(E)$. Then,

$$P = \left(\sum_i a_i P_i \right)(\cdot \mid E) = \frac{\sum_i a_i P_i(\cdot \cap E)}{\sum_i a_i P_i(E)} = \sum_i \frac{a_i P_i(E)}{N} \frac{P_i(\cdot \cap E)}{P_i(E)}$$

$$= \sum_i b_i \frac{P_i(\cdot \cap E)}{P_i(E)} = \sum_i b_i P_i(\cdot \mid E) \in \mathrm{co}\left(\mathbb{P}^E\right).$$

Hence, $\mathrm{co}(\mathbb{P})^E \subseteq \mathrm{co}\left(\mathbb{P}^E\right)$, and the proof is complete. $\qquad\square$

Using Lemmas 1 and 2, if (I) holds, then for all $E \in \mathscr{B}$

$$d(\mathrm{co}(\mathbb{P})) = d(\mathbb{P}) < d\left(\mathbb{P}^E\right) = d\left(\mathrm{co}\left(\mathbb{P}^E\right)\right) = d\left(\mathrm{co}\left(\mathbb{P}\right)^E\right).$$

Hence, (III) holds. And if (III) holds, then for all $E \in \mathscr{B}$

$$d(\mathbb{P}) = d\left(\mathrm{co}\left(\mathbb{P}\right)\right) < d\left(\mathrm{co}\left(\mathbb{P}\right)^E\right) = d\left(\mathrm{co}\left(\mathbb{P}^E\right)\right) = d\left(\mathbb{P}^E\right).$$

Hence, (I) holds. This shows that (I) and (III) are equivalent. $\qquad\square$

References

Arló-Costa, H. and J. Helzner (2010). Ambiguity aversion: The explanatory power of indeterminate probabilities. *Synthese 172*(1), 37–55.

Augustin, T., F. P. Coolen, G. de Cooman, and M. C. Troffaes (2014). *Introduction to Imprecise Probabilities*. West Sussex: John Wiley & Sons.

Berger, J. O. (1994). An overview of robust bayesian analysis. *Test 3*(1), 5–124.

Billingsley, P. (2008). *Probability and Measure*. John Wiley & Sons.

Blackwell, D. and L. E. Dubins (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics 33*(3), 882–886.

Bronevich, A. and G. J. Klir (2010). Measures of uncertainty for imprecise probabilities: An axiomatic approach. *International Journal of Approximate Reasoning 51*(4), 365–390.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Edwards, W., H. Lindman, and L. J. Savage (1963). Bayesian statistical inference for psychological research. *Psychological Review 70*(3), 193–242.

Elkin, L. and G. Wheeler (2018). Resolving peer disagreements through imprecise probabilities. *Noûs 52*(2), 260–278.

Gaifman, H. and M. Snir (1982). Probabilities over rich languages, testing and randomness. *The Journal of Symbolic Logic 47*(03), 495–548.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science 17*(4), 319–321.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. U of Minnesota Press.

Herron, T., T. Seidenfeld, and L. Wasserman (1994). The extent of dilation of sets of probabilities and the asymptotics of robust bayesian inference. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1994, pp. 250–259. Philosophy of Science Association.

Herron, T., T. Seidenfeld, and L. Wasserman (1997). Divisive conditioning: Further results on dilation. *Philosophy of Science 64*(3), 411–444.

Huttegger, S. M. (2015). Merging of opinions and probability kinematics. *The Review of Symbolic Logic 8*(04), 611–648.

Jeffrey, R. (2004). *Subjective Probability: The Real Thing*. Cambridge University Press.

Kadane, J. B., M. Schervish, and T. Seidenfeld (2008). Is ignorance bliss? *The Journal of Philosophy 105*(1), 5–36.

Kadane, J. B., M. J. Schervish, and T. Seidenfeld (1996). Reasoning to a foregone conclusion. *Journal of the American Statistical Association 91*(435), 1228–1235.

Kalai, E. and E. Lehrer (1994). Weak and strong merging of opinions. *Journal of Mathematical Economics 23*(1), 73–86.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge University Press.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy 71*(13), 391–418.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.

Levi, I. (1982). Conflict and social agency. *The Journal of Philosophy 79*(5), 231–247.

Levi, I. (1985). Consensus as shared agreement and outcome of inquiry. *Synthese 62*(1), pp. 3–11.

Levi, I. (1986). The paradoxes of Allais and Ellsberg. *Economics and Philosophy 2*(1), 23–53.

Levi, I. (1990). Pareto unanimity and consensus. *The Journal of Philosophy 87*(9), 481–492.

Levi, I. (2009). Why indeterminate probability is rational. *Journal of Applied Logic 7*(4), 364–376.

Nielsen, M. and R. T. Stewart (2019). Counterexamples to some characterizations of dilation. *Erkenntnis https://doi.org/10.1007/s10670-019-00145-y*, 1–12.

Nielsen, M. and R. T. Stewart (2020). Persistent disagreement and polarization in a Bayesian setting. *The British Journal for the Philosophy of Science*, Forthcoming.

Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis 79*(6), 1305–1342.

Pedersen, A. P. and G. Wheeler (2015). Dilation, disintegrations, and delayed decisions. In *ISIPTA '15: Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications.* Aracne Editrice.

Savage, L. (1972, originally published in 1954). *The Foundations of Statistics.* New York: John Wiley and Sons.

Schervish, M. and T. Seidenfeld (1990). An approach to consensus and certainty with increasing evidence. *Journal of Statistical Planning and Inference 25*(3), 401–414.

Schervish, M. J., T. Seidenfeld, and J. B. Kadane (1984). The extent of non-conglomerability of finitely additive probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 66*(2), 205–226.

Schervish, M. J., T. Seidenfeld, and J. B. Kadane (2017). Non-conglomerability for countably additive measures that are not κ-additive. *Review of Symbolic Logic 10*(2), 284–300.

Seidenfeld, T. (1993). Outline of a theory of partially ordered preferences. *Philosophical Topics 21*(1), 173–189.

Seidenfeld, T., J. B. Kadane, and M. J. Schervish (1989). On the shared preferences of two bayesian decision makers. *The Journal of Philosophy 86*(5), 225–244.

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2010). Coherent choice functions under uncertainty. *Synthese 172*(1), 157–176.

Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics 21*(3), 1139–1154.

Stewart, R. T. and M. Nielsen (2019). Another approach to consensus and maximally informed opinions with increasing evidence. *Philosophy of Science 86*(2), 1–19.

Stewart, R. T. and I. Ojea Quintana (2018). Probabilistic opinion pooling with imprecise probabilities. *Journal of Philosophical Logic 47*(1), 17–45.

Topey, B. (2012). Coin flips, credences and the reflection principle. *Analysis 72*(3), 478–488.

Wagner, C. (2002). Probability kinematics and commutativity. *Philosophy of Science 69*(2), 266–278.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities.* Chapman and Hall London.

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning 24*(2-3), 125–148.

Wasserman, L. and T. Seidenfeld (1994). The dilation phenomenon in robust bayesian inference. *Journal of Statistical Planning and Inference 40*(2), 345–356.

White, R. (2010). Evidential symmetry and mushy credence. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Volume 3, pp. 161–186. Oxford University Press New York.