Another Approach to Consensus and Maximally Informed Opinions with Increasing Evidence

Rush T. Stewart and Michael Nielsen*†

Merging of opinions results underwrite Bayesian rejoinders to complaints about the subjective nature of personal probability. Such results establish that sufficiently similar priors achieve consensus in the long run when fed the same increasing stream of evidence. Here, we establish a merging result for sets of probability measures updated by Jeffrey conditioning. This generalizes a number of different merging results in the literature. We also show that such sets converge to a shared, maximally informed opinion. Finally, we demonstrate the philosophical significance of our study by detailing applications to the topics of dynamic coherence, imprecise probabilities, and opinion pooling.

1. Introduction. Merging of opinions results (e.g., Blackwell and Dubins 1962; Gaifman and Snir 1982; Schervish and Seidenfeld 1990; Kalai and Lehrer 1994; Huttegger 2015b) have underwritten Bayesian rejoinders to complaints about the subjective nature of personal probability (e.g., Savage 1954/1972; Schervish and Seidenfeld 1990; Earman 1992). Such results establish that sufficiently similar priors achieve consensus in the long run when fed the same increasing stream of evidence. Initial subjectivity, the line goes, is of mere transient significance, giving way to intersubjective agreement even-

Received May 2017; revised May 2018.

*To contact the authors, please write to: Rush T. Stewart, Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich; e-mail: rush.stewart@lrz.uni-muenchen.de. Michael Nielsen, Department of Philosophy, Columbia University; e-mail: mn2683@columbia.edu.

†We would like to thank Jean Baccelli, Jessica Collins, Robby Finley, Arthur Heller, Simon Huttegger, Isaac Levi, Ignacio Ojea Quintana, Teddy Seidenfeld, Reuben Stern, Greg Wheeler, an audience at the Institute of History and Philosophy of Science and Technology in Paris, and two anonymous referees for generous feedback that helped us to improve the article.

Philosophy of Science, 86 (April 2019) pp. 236–254. 0031-8248/2019/8602-0003$10.00
Copyright 2019 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

tually. Schervish and Seidenfeld (1990) show that Blackwell and Dubins's classic result (1962) can be extended to certain sets of probability functions, while Huttegger (2015b) provides sufficient conditions for merging of opinions for Jeffrey conditioning, a generalization of Bayesian conditionalization. We establish that Huttegger's merging result for Jeffrey conditioning can in turn be extended to sets of probability functions in analogy to Schervish and Seidenfeld's extension (sec. 7). We also extend Huttegger's convergence result (to a "maximally informed opinion") for Jeffrey conditioning to convergence of a set of probabilities to a shared maximally informed opinion (sec. 8). Convergence to a maximally informed opinion is a (weak) Jeffrey conditioning analogue of Bayesian "convergence to the truth" for conditional probabilities.

There are a number of motivations for considering merging and convergence for sets of probabilities. For example, according to proponents of imprecise probabilities (IP), rationality does not always demand numerically precise probability judgments of an agent—to say nothing of the descriptive adequacy of such precise judgments (e.g., Levi 1974; Walley 1991). In this context, merging and convergence results show that a single agent with a certain sort of imprecise credal state fully expects to uniformly strengthen her point of view toward a consensus among the set (merging) and for the consensus point of view to stabilize (convergence). In effect, evidence reduces an agent's imprecise credal state toward a precise probability, at least in certain cases. In section 9.2, we discuss how our results establish that "certain evidence" is unnecessary for the reduction of an imprecise credal state toward a stable, precise probability judgment. Another motivation comes from considering these points in a social setting. In such a setting, sets represent beliefs in some community. In section 9.3, we detail an application of our study to probabilistic opinion pooling with IP (Elkin and Wheeler 2018; Stewart and Ojea Quintana 2018b).

We also discuss Schervish and Seidenfeld's (1990) claim that such results for sets of probabilities allow us to relax so-called dynamic coherence to a large extent. In the case of Bayesian updating, for certain asymptotic properties like convergence to certainty and long-run consensus, it is enough that posteriors take any value in a particular set. In principle, our propositions 1 and 2 establish that a similar point can be made about relaxing dynamic coherence for Jeffrey conditioning. However, we discuss some reservations about the precise significance of this point in section 9.1.

2. Preliminaries. Let Ω be a sample space, a set of elementary events or possible worlds. We let $\mathfrak{F}$ denote a σ-algebra of subsets of Ω, that is, a nonempty set of subsets of Ω closed under complementation and countable unions. Elements of $\mathfrak{F}$ can be interpreted as events or propositions. For example, Ω could be the set of all infinite sequences of tosses of a coin, and $\mathfrak{F}$ would be the set of relevant coin-tossing events. Included in $\mathfrak{F}$ would be proposi-

https://doi.org/10.1086/701954 Published online by Cambridge University Press

tions describing finite initial segments of a sequence like the first flip landing heads, as well as propositions describing the limiting behavior of the sequence such as $\lim_{n\to\infty} H_n/n = 1/4$, where H_n is the total number of heads on the first n flips. Throughout the article uppercase blackboard letters like $\mathbb{P}$ and $\mathbb{Q}$ denote (countably additive) probability measures on $(\Omega, \mathfrak{F})$.[1] An event $A \in \mathfrak{F}$ is said to occur *almost surely* with respect to $\mathbb{P}$, or a.s. $(\mathbb{P})$, if $\mathbb{P}[A] = 1$.

Let $\mathfrak{E}_1, \mathfrak{E}_2, \ldots$ be an infinite sequence of (finite) partitions of Ω. We suppose that, for any n, $\mathfrak{E}_{n+1}$ is a refinement of $\mathfrak{E}_n$; that is, every element of $\mathfrak{E}_{n+1}$ is a subset of an element of $\mathfrak{E}_n$. For any n, the partition $\mathfrak{E}_n$ can be thought of as the possible information an agent might receive about the actual world $\omega \in \Omega$. A single flip of a coin determines a binary partition of the set of all infinite sequences of tosses: the coin lands heads or the coin lands tails.

Let $\mathfrak{F}_n$ be the algebra generated by $\mathfrak{E}_n$. Apart from $\varnothing$ and Ω, $\mathfrak{F}_n$ contains all the unions of elements of $\mathfrak{E}_n$. In the coin-tossing example, the algebra $\mathfrak{F}_n$ can be thought of as the set of all coin-tossing events up to stage n. The refinement assumption on $\mathfrak{E}_1, \mathfrak{E}_2, \ldots$ implies that $\mathfrak{F}_n \subseteq \mathfrak{F}_{n+1}, n = 1, 2, \ldots$. We say that the sequence $\{\mathfrak{F}_n : n \in \mathbb{N}\}$ of σ-algebras is a *filtration*. We assume that this filtration increases to the background σ-algebra $\mathfrak{F}$ (i.e.,$\sigma(\cup_{n\geq 0}\mathfrak{F}_n) = \mathfrak{F}$), so that $\mathfrak{F}$ is generated by a countable sequence of subsets of Ω. The informal idea behind the filtration structure is that information is always increasing and eventually captures all propositions of interest.

For any events $A, E \in \mathfrak{F}$ with $\mathbb{P}[E] > 0$, we use the standard ratio definition of conditional probability given an event:

$$\mathbb{P}[A|E] = \frac{\mathbb{P}[A \cap E]}{\mathbb{P}[E]}. \tag{1}$$

For a sub-σ-algebra $\mathfrak{F}_n$ of $\mathfrak{F}$, we will also use the standard definition of conditional probability given $\mathfrak{F}_n$: $\mathbb{P}[A|\mathfrak{F}_n]$ is a $\mathfrak{F}_n$-measurable function that satisfies

$$\mathbb{P}[A \cap E] = \int_E \mathbb{P}[A|\mathfrak{F}_n](\omega)\mathbb{P}[d\omega] \tag{2}$$

for all $E \in \mathfrak{F}_n$.[2] If we think of ω as the "actual world" regarding coin tosses, then, after observing n tosses, $\mathfrak{F}_n$ is the collection of all propositions that the agent can distinguish as true or false in the actual world. So, just as the

1. The role of countable additivity in a normative theory of probability judgment is a contentious issue. Its mathematical role in all of the results discussed here is not: it is presupposed.

2. For any sub-σ-algebra $\mathfrak{G}$, the existence of the conditional probability $\mathbb{P}[A|\mathfrak{G}]$ is guaranteed by the Radon-Nikodym theorem. The uniqueness of $\mathbb{P}[A|\mathfrak{G}]$ is only almost sure. This means that, in general, there are many measurable functions that satisfy eq. (2) and differ from each other on a set of $\mathbb{P}$-measure 0. To mark this fact, we say that there are different

https://doi.org/10.1086/701954 Published online by Cambridge University Press

conditional probability of A given an event E can be interpreted as the probability of A on the supposition that E is true, so the conditional probability of A given $\mathfrak{F}_n$ can be interpreted as the probability of A in light of the information provided by $\mathfrak{F}_n$.[3] In the current framework, conditional probabilities given sub-σ-algebras can be reduced (a.s.) to conditional probabilities given events; the former merely provide a convenient tool for working around conditioning on null events without adopting extra conventions or assumptions. Since each $\mathfrak{F}_n$ is generated by a finite partition $\mathfrak{E}_n$, for almost every $\omega \in \Omega$ we have

$$\mathbb{P}[A|\mathfrak{F}_n](\omega) = \mathbb{P}(A|E_n(\omega)), \tag{3}$$

where $E_n(\omega)$ is the unique cell of $\mathfrak{E}_n$ that contains ω. If $\mathbb{P}(E_n(\omega)) = 0$, then the left-hand side of (3) can be defined arbitrarily.

3. Merging of Opinions. What does it mean to say that two opinions merge, or that $\mathbb{P}$ and $\mathbb{Q}$ agree in the limit? Following Blackwell and Dubins (1962), we adopt the *total variation distanced* as a measure of the distance between $\mathbb{P}$ and $\mathbb{Q}$.

$$d(\mathbb{P}, \mathbb{Q}) = \sup_{A \in \mathfrak{F}} |\mathbb{P}[A] - \mathbb{Q}[A]|.$$

So if $\mathbb{P}$ and $\mathbb{Q}$ are within ε according to the metric d, then they are within ε for every event $A \in \mathfrak{F}$. Now we can formulate a natural sense in which two sequences $\{p_n\}$ and $\{q_n\}$ of probability measures might become (and stay) close. We say that $\{p_n\}$ and $\{q_n\}$ *merge* if

$$d(p_n, q_n) \to 0$$

as $n \to \infty$. If $\{p_n\}$ and $\{q_n\}$ are updates (not necessarily Bayesian) of $\mathbb{P}$ and $\mathbb{Q}$ respectively, we say that $\mathbb{P}$ and $\mathbb{Q}$ merge if $\{p_n\}$ and $\{q_n\}$ do.

The main result of Blackwell and Dubins (1962) concerns the merging of conditional probabilities. If learning goes by conditionalization, and the priors $\mathbb{P}$ and $\mathbb{Q}$ do not disagree too drastically, then the Blackwell-Dubins result says $\mathbb{P}$ and $\mathbb{Q}$ must assign probability 1 to merging. To make this precise, we

versions of $\mathbb{P}[A|\mathfrak{G}]$. Unlike conditional probabilities given events, there may not exist versions of a conditional probability given a sub-σ-algebra such that $\mathbb{P}[\cdot|\mathfrak{G}](\omega)$ is a probability measure on $(\Omega, \mathfrak{F})$ for all $\omega \in \Omega$. In this case, we say that $\mathbb{P}[\cdot|\mathfrak{G}]$ is *irregular*. We discuss regularity more below in connection with the Blackwell-Dubins merging of opinions theorem.

3. As pointed out by Billingsley (2008, 438), this heuristic explanation of conditioning on a sub-σ-algebra breaks down when conditional probabilities are not proper (see also Blackwell and Dubins 1975; Seidenfeld 2001).

https://doi.org/10.1086/701954 Published online by Cambridge University Press

say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$, denoted $\mathbb{Q} \ll \mathbb{P}$, if for all $A \in \mathfrak{F}$

$$\mathbb{P}[A] = 0 \Rightarrow \mathbb{Q}[A] = 0.$$

So $\mathbb{Q}$ agrees with $\mathbb{P}$ about events bearing probability 0. If $\mathbb{P}$ also agrees with $\mathbb{Q}$ about probability 0 events, so that $\mathbb{P} \ll \mathbb{Q}$, we say that $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent* or *mutually absolutely continuous*. In that case, we have for all $A \in \mathfrak{F}$

$$\mathbb{P}[A] = 0 \Leftrightarrow \mathbb{Q}[A] = 0.$$

It turns out that absolute continuity is sufficient for the merging of conditional probabilities.[4] We write $\mathbb{P}_{\mathfrak{F}_n} = \mathbb{P}[\cdot|\mathfrak{F}_n]$ and $\mathbb{Q}_{\mathfrak{F}_n} = \mathbb{Q}[\cdot|\mathfrak{F}_n]$.[5]

> **Theorem 1.** If $\mathbb{Q} \ll \mathbb{P}$, then $d(\mathbb{P}_{\mathfrak{F}_n}, \mathbb{Q}_{\mathfrak{F}_n}) \to 0$ a.s. $(\mathbb{Q})$ as $n \to \infty$. (Blackwell and Dubins 1962, 883)

With $\mathbb{Q}$-probability 1, $\mathbb{P}$ merges to $\mathbb{Q}$ when $\mathbb{Q} \ll \mathbb{P}$. If $\mathbb{P} \ll \mathbb{Q}$ also, we have merging from the perspectives of both $\mathbb{P}$ and $\mathbb{Q}$. The almost surely qualifications are needed in the statement of theorem 1 because the conditional probabilities $\mathbb{P}_{\mathfrak{F}_n}$ and $\mathbb{Q}_{\mathfrak{F}_n}$ in this general setting are random objects; that is, they depend on $\omega \in \Omega$ (see n. 2 for details).

4. Merging of Opinions for Sets of Probabilities. Schervish and Seidenfeld (1990) provide sufficient conditions to secure merging of opinions for *sets* of probability functions. A set of probability functions C is called *convex* if $\mathbb{P}, \mathbb{Q} \in C$ implies $\alpha\mathbb{P} + (1 - \alpha)\mathbb{Q} \in C$, where $\alpha \in [0, 1]$. That is, for any two elements of C, C includes all convex combinations of those elements. We call $\mathbb{P} \in C$ an *extreme point* of C if $\mathbb{P} = \alpha\mathbb{Q} + (1 - \alpha)\mathbb{R}$ with $\mathbb{Q}$, $\mathbb{R} \in C$, and $\alpha \in [0, 1]$ implies $\mathbb{Q} = \mathbb{P}$ or $\mathbb{R} = \mathbb{P}$. The following result is a consequence of theorem 1.

> **Theorem 2.** Let C be a closed, convex set of probability functions, all mutually absolutely continuous, and generated by finitely many extreme points. Then, almost surely ($\mathbb{P} \in C$), the conditional probabilities of elements of C merge uniformly; that is, $\sup_{\mathbb{P},\mathbb{Q}\in C} d(\mathbb{P}_{\mathfrak{F}_n}(\omega), \mathbb{Q}_{\mathfrak{F}_n}(\omega)) \to 0$ as $n \to \infty$ for almost all $\omega \in \Omega$. (Schervish and Seidenfeld 1990, corollary 1)

4. Kalai and Lehrer (1994, theorem 2) show that absolute continuity is not merely a sufficient condition for merging of conditional probabilities; it is also a necessary one.

5. Apart from absolute continuity, the Blackwell-Dubins theorem also requires that the conditional probabilities $\mathbb{P}_{\mathfrak{F}_n}$ and $\mathbb{Q}_{\mathfrak{F}_n}$ be *regular* (or what Blackwell and Dubins call *predictive*): $\mathbb{P}_{\mathfrak{F}_n}(\omega)$ and $\mathbb{Q}_{\mathfrak{F}_n}(\omega)$ are probability measures on $(\Omega, \mathfrak{F})$ for each $\omega \in \Omega$. It follows from the fact that each $\mathfrak{F}_n$ is generated by a finite partition that $\mathbb{P}_{\mathfrak{F}_n}$ and $\mathbb{Q}_{\mathfrak{F}_n}$ have regular versions. We assume throughout that all conditional probabilities are regular. For examples of conditional probabilities that are irregular, see Seidenfeld (2001) and Billingsley (2008, exercise 33.11).

https://doi.org/10.1086/701954 Published online by Cambridge University Press

Another way of stating the conclusion of theorem 2 is that for almost all ω and all $\varepsilon > 0$ there is some m such that for all $n \geq m$ and all $\mathbb{P}$, $\mathbb{Q} \in C$, $d(\mathbb{P}_{\mathfrak{F}_n}(\omega), \mathbb{Q}_{\mathfrak{F}_n}(\omega)) < \varepsilon$. That the same m works for all $\mathbb{P}$ and $\mathbb{Q}$ is what it means for merging to be *uniform*.

Here is the gist of their proof. Since C is generated by finitely many extreme points, $\mathbb{P}^1, \ldots, \mathbb{P}^k$, the (total variation) distance between the conditional probabilities of any of the extreme points of C is bounded by the maximum distance for pairs of conditional probabilities of extreme points. Moreover, that maximum is a bound for the distance between the conditional probabilities of any points in the whole set. To see this, note two things. First, $d(\mathbb{P}, \mathbb{Q})$ bounds the distances $d(\mathbb{P}, \mathbb{R})$ and $d(\mathbb{R}, \mathbb{Q})$, where $\mathbb{R}$ is a convex combination of $\mathbb{P}$ and $\mathbb{Q}$. Second, at any stage n, the conditional probabilities of points in C can be written as convex combinations of the conditional probabilities of the extreme points; that is, there exist $\alpha_1, \ldots, \alpha_k$ with $\alpha_i \in [0, 1]$ and $\Sigma_{i=1}^k \alpha_i = 1$ such that $\mathbb{R}_{\mathfrak{F}_n} = \Sigma_{i=1}^k \alpha_i \mathbb{P}^i_{\mathfrak{F}_n}$ for all $\mathbb{R} \in C$ and all n. Theorem 2 is stronger than it may at first appear. Notice that any set of probabilities that is included in a closed, convex set generated by finitely many extreme points merges uniformly by this result.

It is not easy to relax the assumption in theorem 2 that C is convex and generated by finitely many extreme points. Schervish and Seidenfeld (1990) demonstrate this with several other results and examples. For instance, if C is d-compact, an assumption that is weaker than the assumptions of theorem 2, the conditional probabilities of elements of C may not exhibit almost sure uniform merging. However, Schervish and Seidenfeld show that weaker modes of merging can be achieved under weaker assumptions about C, such as d-compactness (see their corollaries 2 and 3). It is worth noting that any negative result for Bayesian conditionalization and merging carries over to Jeffrey conditioning, a generalization of conditionalization that we introduce and study below. An immediate consequence of Schervish and Seidenfeld's results, then, is that Jeffrey updates of a d-compact set of probability measures may not a.s. merge uniformly.

There are a number of uses to which results such as theorem 2 might be put. We delay explaining some of the significance of theorem 2 and propositions 1 and 2 (to come) until section 9.

5. Jeffrey Conditioning. Bayesian conditionalization is a putative diachronic norm that specifies how probabilistic learning takes place.[6] A basic assumption of conditionalization is that it is propositions (or events) that are learned.

6. In this essay, we will not quarrel with the view that conditionalization and Jeffrey conditioning are genuine diachronic norms (outside of indicating how they can be relaxed while still achieving merging and convergence). However, both authors regard diachronic "learning" norms with a good deal of suspicion. We recognize that this is likely at odds with most philosophical writing on probability.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

In other words, learning experiences can always be represented by some $E \in \mathfrak{F}$. When an agent learns E, conditionalization says that her *posterior* probability $\mathbb{P}_E[A]$ for an event $A \in \mathfrak{F}$ should be equal to her prior conditional probability of A given E (provided, of course, that this is well defined):

$$\mathbb{P}_E[A] = \mathbb{P}[A|E].$$

It is clear from (1) that conditionalization requires that E have posterior probability 1. Jeffrey conditioning relaxes both of the fundamental features of conditionalization: it does not assume that learning experiences are always represented by propositions, and it does not require that learning involves assigning propositions probability 1.

Proponents of Jeffrey conditioning want to allow for "uncertain learning." Uncertain learning induces a change in the probabilities P_n assigned to the members of the partition $\mathfrak{E}_n$. Jeffrey conditioning applies just in case the change over $\mathfrak{E}_n$ is *rigid*:

$$\mathbb{P}_n[A|E] = \mathbb{P}_{n-1}[A|E], \text{ for all } E \in \mathfrak{E}_n \text{ and all } A \in \mathfrak{F}. \tag{4}$$

The rigidity condition (4) says that the update from $\mathbb{P}_{n-1}$ to $\mathbb{P}_n$ does not change conditional probabilities given members of the partition $\mathfrak{E}_n$. This is equivalent to the requirement that for all members E of the partition $\mathfrak{E}_n$ and all subsets A, B of E, the update from $\mathbb{P}_{n-1}$ to $\mathbb{P}_n$ does not change the ratio of the probabilities of A and B. The law of total probability and (4) yield the familiar Jeffrey conditioning equation, which extends P_n from $\mathfrak{E}_n$ to the entire algebra $\mathfrak{F}$:

$$\mathbb{P}_n[A] = \sum_{E \in \mathfrak{E}_n} \mathbb{P}_{n-1}[A|E] P_n[E], \text{ for all } E \in \mathfrak{E}_n \text{ and all } A \in \mathfrak{F}. \tag{5}$$

If $\mathfrak{E}_n = \{E, E^c\}$ and $P_n[E] = 1$, then Jeffrey conditioning reduces to standard conditionalization; that is, $\mathbb{P}_n[A] = \mathbb{P}_{n-1}[A|E]$ for all $A \in \mathfrak{F}$.[7] Equation (5) simplifies in another way under our assumption that the sequence of partitions $\mathfrak{E}_1, \ldots, \mathfrak{E}_n$ is such that $\mathfrak{E}_{i+1}$ refines $\mathfrak{E}_i$ for all $1 \leq i \leq n - 1$. In this case, after $n - 1$ applications of equation (4) in equation (5), we have

$$\mathbb{P}_n[A] = \sum_{E \in \mathfrak{E}_n} \mathbb{P}[A|E] P_n[E], \text{ for all } E \in \mathfrak{E}_n \text{ and all } A \in \mathfrak{F}. \tag{6}$$

Equation (6) shows that, in our framework, posterior probabilities $\mathbb{P}_n$ are determined by their values on members of the partition $\mathfrak{E}_n$ and prior ($\mathbb{P}$) conditional probabilities given members of $\mathfrak{E}_n$. Unless otherwise stated, we will be assuming that the events $E \in \mathfrak{E}_n$ have positive prior probability so that the right-hand side of (6) is well defined.

7. This, of course, requires that we adopt some conventions for conditioning on null events; otherwise, the left-hand side of eq. (4) may be undefined.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

It is worth pointing out that, in certain situations, Jeffrey conditioning can be represented as standard Bayesian conditionalization in a richer algebra. The cases in which such so-called *superconditioning* is possible are characterized by Diaconis and Zabell's superconditioning criterion (1982, theorem 2.1). But the superconditioning criterion is not trivial; it fails to hold in many cases. Hence, Huttegger's merging result for Jeffrey conditioning in the following section and our merging result for sets of Jeffrey updates in section 7 genuinely generalize Blackwell and Dubins's and Schervish and Seidenfeld's results, respectively. (For other reservations about reducing Jeffrey conditioning to Bayesian conditionalization in the setting of merging of opinions, see Huttegger's [2015b, 630–31] discussion.)

6. Merging of Opinions for Jeffrey Conditioning. Huttegger proves an analogue of Blackwell and Dubins's merging result for Jeffrey conditioning. Just as we considered random conditional probabilities in the case of conditionalization, we now treat posteriors over the learning partition as random. By random probability, we mean $P_n[\cdot]$ is not a determinate number but rather a (measurable) function on Ω such that $P_n[\cdot](\omega)$ is a probability measure on $\mathfrak{E}_n$ for each $\omega \in \Omega$. Let $p_n^E = P_n[E] = \mathbb{P}_n[E]$ be a random probability assigned to $E \in \mathfrak{E}_n$. We treat future probabilities as random quantities because, from an agent's present point of view concerning $E \in \mathfrak{E}_n$, the precise value p_n^E is unknown.

Because Jeffrey conditioning is less constrained than Bayesian conditionalization, some additional assumptions figure into Huttegger's theorem. The first additional assumption is that the sequence $\{Q_n\}$ is *uniformly absolutely continuous* with respect to $\mathbb{Q}$. This condition consists of two requirements: (i) $Q_n \ll \mathbb{Q}|_{\mathfrak{F}_n}$ for all n, and (ii) for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all n and all $A \in \mathfrak{F}_n$,

$$\mathbb{Q}[A] < \delta \Rightarrow Q_n[A] < \varepsilon.$$

Requirement i is the usual constraint on Bayesian learning that events that are null for the prior $\mathbb{Q}$ remain null for the posterior $\mathbb{Q}_n$. The additional uniformity requirement ii ensures that this relation holds as we pass to the limit (Huttegger 2015b, 623).[8]

8. Here is another way of thinking about the uniform absolute continuity requirement. First, if $Q_n \ll \mathbb{Q}|_{\mathfrak{F}_n}$ for all n, then $\{Q_n\}$ is uniformly absolutely continuous with respect to $\mathbb{Q}$ if and only if the sequence $\{Y_n\} = \{dQ_n/d\mathbb{Q}\}$ of Radon-Nikodym derivatives is uniformly integrable. The "only if" direction of this statement is shown by Huttegger (2015b, lemma 12.1), and the "if" direction is not difficult to prove. The uniform integrability of $\{Y_n\}$ in turn implies that $\{Y_n\}$ is uniformly bounded in expectation, i.e., $\sup_n \mathbb{E}[Y_n] < \infty$ (again, this is a standard result). Therefore, if we think of the derivatives Y_n as representing the "rate" at which the posteriors Q_n are changing with respect to the prior $\mathbb{Q}$, then Huttegger's uniform absolute continuity condition demands that this rate is not expected to diverge to ∞.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

The other assumption demands a certain sort of stability of probability judgments as the agent updates over increasingly fine partitions. While we treat posterior probabilities as random, condition (M′) places significant constraints on sequences of probability judgments. For all n and all $E \in \mathfrak{E}_n$,

$$\int_G p^E_{m+1} d\mathbb{P} = \int_G p^E_m d\mathbb{P} \text{ for all } m \geq n \text{ and all } G \in \mathfrak{F}_m M' \tag{M′}$$

Property (M′) is the *martingale* condition, and it requires that future probabilities are equal to present probabilities on average. For all n and all $E \in \mathfrak{E}_n$, the sequence $p^E_n, p^E_{n+1}, \ldots$ must form a martingale. The martingale condition has been defended as an essential feature of rational learning experiences (Huttegger 2013, 2015b). In the case of Bayesian conditionalization, future probabilities are fixed at 1 for those events learned. This, of course, is not generally the case for Jeffrey conditioning.

Theorem 3 is Huttegger's merging result for Jeffrey conditioning.

> **Theorem 3.** Suppose that P_n and Q_n, $n = 1, 2, \ldots$ are random sequences of probability measures on $(\Omega, \mathfrak{F}_1)$, $(\Omega, \mathfrak{F}_2)$, . . . with $Q_n = P_n$, that Q_n, $n = 1, 2, \ldots$ is uniformly absolutely continuous with respect to $\mathbb{Q}$ a.s. $(\mathbb{Q})$, and that $\mathbb{Q} \ll \mathbb{P}$. If (M′) holds for Q_n, $n = 1, 2, \ldots$, then $d(\mathbb{P}_n, \mathbb{Q}_n) \to 0$ as $n \to \infty$ a.s. $(\mathbb{Q})$. (Huttegger 2015b, theorem 9.2)

7. Merging of Opinions for Jeffrey Conditioning and Sets of Probabilities. We are now in a position to turn to the contributions of the current article. Our first result extends Huttegger's theorem for merging of opinions via Jeffrey conditioning to sets of probabilities, an analogue of theorem 2.

> **Proposition 1.** Let C be a closed, convex set of probability measures, all mutually absolutely continuous and generated by finitely many extreme points. Suppose that a.s. each pair of probabilities $\mathbb{P}, \mathbb{Q} \in C$ satisfies the conditions of theorem 3. Then a.s. $(\mathbb{P} \in C)$ the elements of C merge uniformly.

A proof is included in the appendix. As Schervish and Seidenfeld (1990) indicate, theorem 2 is a simple corollary of theorem 1 given that point-by-point conditionalization preserves the convexity of a set. We find proposition 1 especially interesting in light of the fact that, unlike Bayesian conditionalization, Jeffrey conditioning does not generally preserve the convexity of the initial set—even for convex sets of priors generated by finitely many extreme points. We provide a demonstration in the appendix (proposition 3).

8. Maximally Informed Opinions. Another consequence of Doob's martingale convergence theorem is the so-called convergence to truth or con-

https://doi.org/10.1086/701954 Published online by Cambridge University Press

vergence to certainty for conditional probabilities, captured by the following statement:

$$\text{For all } A \in \mathfrak{F}, \lim_{n\to\infty} \mathbb{P}_{\mathfrak{F}_n}[A] = 1_A \text{ a.s. } (\mathbb{P}).$$

Put another way, coherence requires assigning probability 1 to converging to "the truth" for any event $A \in \mathfrak{F}$ whose truth value is determined by observations. One way of interpreting such convergence is that "it shows that evidence triumphs over prior opinions under the appropriate circumstances" (Huttegger 2015a, 590). Two mutually absolutely continuous priors assign probability 1 not just to merging but to being certain of the same observational hypotheses in the limit.[9]

Convergence to certainty does not obtain in general for Jeffrey conditioning. However, Huttegger shows that Jeffrey conditioning does lead probabilities to settle down to some limit if we assume (M′), just not necessarily to 0 or 1.[10] In particular, Huttegger's (2015b) theorem 9.1 states that sequences of a.s. uniformly absolutely continuous Jeffrey updates that satisfy (M′) converge setwise a.s.; that is, for all $A \in \mathfrak{F}$, $\mathbb{P}_\infty[A] \coloneqq \lim_{n\to\infty} \mathbb{P}_n[A]$ exists a.s. Skyrms calls this convergence to a "maximally informed opinion" (1996).

Convergence to maximally informed opinions can be extended to sets of probabilities. For all $A \in \mathfrak{F}$, all probabilities in the set C converge to the same limit $\mathbb{P}_\infty[A]$ a.s. To see this, consider that for all $A \in \mathfrak{F}$, by the result of Huttegger just mentioned, all elements in C a.s. are converging to a limit. Also, all the elements of C are merging (theorem 3). If two elements $\mathbb{P}, \mathbb{Q} \in C$ converged to different limits (with positive probability) for some $A \in \mathfrak{F}$, then there would be some ε and some m such that $d(\mathbb{P}_n, \mathbb{Q}_n) > \varepsilon$ for any $n \geq m$, which is inconsistent with merging. In fact, a slightly stronger observation can be demonstrated using our proposition 1. As indicated in the introduction, dynamic coherence can be relaxed. For convergence to a maximally informed opinion, it suffices that probabilities take their values in the set of Jeffrey updates of elements of C.

> **Proposition 2.** Let C be as in proposition 1, and let $A \in \mathfrak{F}$. If $f_n[A] \in \{\mathbb{P}_n[A] : \mathbb{P} \in C\}$, then the sequence $f_n[A]$ converges to the maximally informed opinion $\mathbb{P}_\infty[A]$ a.s. ($\mathbb{P} \in C$).

We omit the proof. Proposition 2 is the Jeffrey conditioning analogue of Schervish and Seidenfeld's corollary 4 (1990, 410).

9. While Cisewski et al. (2018), e.g., refer to such convergence results as "desirable" and "laudable," others are less enthusiastic (e.g., Earman 1992; Kelly 1996; Belot 2013).

10. For a discussion of this point, see Huttegger (2015b, secs. 7 and 9).

https://doi.org/10.1086/701954 Published online by Cambridge University Press

9. Philosophical Significance. Standard Bayesian learning proceeds by conditionalization, which requires that the event or proposition learned be assigned posterior probability 1. As fans of Jeffrey conditioning sometimes put the point, "learning experiences need not be like that at all" (Huttegger 2015b, 613). Some learning experiences do not lead to certainty. Here, the standard Bayesian framework cannot help. If there are "uncertain learning experiences," Jeffrey conditioning, and the extension of learning rules to accommodate "uncertain evidence" more generally, are topics of considerable philosophical interest. Similarly, as fans of IP often point out, not all uncertainty is reducible to numerically precise probabilities. Here, too, the standard Bayesian framework cannot help. In the presence of "deep uncertainty," sets of probability measures, and generalizations of precise probability models more generally, warrant philosophical attention and scrutiny.

We find it somewhat surprising, then, that with very few exceptions (e.g., Škulj 2006; Stewart and Ojea Quintana 2018a) studies that treat "uncertain learning" do not do so in the context of uncertainty that is not reducible to a numerically precise probability, nor do studies of IP treat uncertain evidence. A very general motivation for our current study is to redress this lacuna in the literature. We turn now to three more specific points of philosophical interest.

9.1. Relaxing Dynamic Coherence. Earlier, we indicated some consequences of merging results for sets of probabilities. First, Schervish and Seidenfeld (1990) claim that results such as theorem 2 and proposition 1 have interesting implications for issues related to dynamic coherence. Precisely what they have in mind about the connection between their results and dynamic coherence remains a bit obscure to us (more on this below). The upshot of their discussion, however, seems to be that to achieve asymptotic consensus, agents need not be dynamically coherent—they need not conditionalize or update by Jeffrey conditioning. As Schervish and Seidenfeld put the point with respect to conditionalizing, "We do not impose a constraint of (full) dynamic coherence. For consensus, it suffices that the agents use conditional probabilities arbitrarily chosen from a class C enveloped by finitely many (mutually absolutely continuous) distributions. Under the conditions of [theorem 2], asymptotic certainty follows from static coherence" (402). Interpreting merging results as claims about agents' unconditional probabilities through time, it is sufficient for achieving asymptotic consensus that each posterior be a conditionalization of some element of C.

By the results of this article, this point can be extended to Jeffrey conditioning. For agents with priors contained within a convex polytope of mutually absolutely continuous probabilities, C, as long as they take posterior probabilities in the set that results from Jeffrey updating the elements of C under the conditions of theorem 3, they will achieve consensus. That is, each agent is free to choose her posterior arbitrarily from the set of Jeffrey updates of

https://doi.org/10.1086/701954 Published online by Cambridge University Press

elements of C without surrendering the assurance of long-run consensus. For the purpose of achieving consensus with elements of C, an agent's posterior need not result from Jeffrey updating *her* prior; it suffices that it be obtained as the result of Jeffrey updating *some* prior in C.

Analogous points hold for convergence. As Schervish and Seidenfeld (1990) point out, a consequence of their results is that convergence to the truth does not require conditionalization. Similarly, it follows from proposition 2 that convergence to a maximally informed opinion does not demand Jeffrey conditioning. Once again, it suffices that an agent's posterior be in the set of Jeffrey updates of elements of C.

While Schervish and Seidenfeld's point that dynamic coherence is not necessary for convergence is certainly correct, and while the results discussed in this article certainly suffice to make this point, the same moral can be gleaned in an even more straightforward way. For example, let $\{f_n : n \in N\}$ be any sequence of "updates" of $\mathbb{Q}$ that satisfies

$$f_n[A](\omega) \in \{x \in \mathbb{R} : |x - \mathbb{Q}_{\mathfrak{F}_n}[A](\omega)| < 1/n\}$$

for all $n \in \mathbb{N}$, all $A \in \mathfrak{F}$, and $\mathbb{Q}$-almost every $\omega \in \Omega$. Then, $f_n[A]$ converges to the truth for all $A \in \mathcal{F}$ a.s. But $\{f_n : n \in \mathbb{N}\}$ need not be a dynamically coherent sequence of updates. In view of simple examples of this kind, it is not clear to us that there is an especially interesting or significant connection between dynamic coherence and the asymptotics of sets of probabilities.

Whether or not one finds such a connection, it is clear that we need not impose full dynamic coherence in order to achieve the desirable epistemic ends of convergence and merging. Implicit in Schervish and Seidenfeld's focus on these sorts of asymptotic properties, however, is the idea that we may not be able to relax dynamic coherence in this way and retain just any good-making property typically associated with it. For example, if one is compelled by dynamic Dutch books, taking arbitrary values in the set as above will not cut it. And Skyrms (1987) has shown that Dutch book arguments can be extended to Jeffrey conditioning. It bears repeating, though, that the status of dynamic Dutch book arguments is a subject of considerable controversy (e.g., Levi 1987; Maher 1992; Skyrms 1993).

9.2. IP and Inquiry for a Single Agent. Second, if an agent has a credal state properly represented by a set of probability measures, theorem 2 and proposition 1 provide sufficient conditions for that indeterminacy to be reduced by evidence in the limit, for both Bayesian conditionalization and Jeffrey conditioning. Standard, precise probability models are often taken as suspect in contexts of "deep" or "Knightean" uncertainty, motivating the appeal to more general models of uncertainty (Ellsberg 1963; Levi 1974; Walley 1991). Sets of probability measures provide one very general representation of uncertainty.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

The elements of such sets can be interpreted as those probabilities that the agent regards as permissible to use in inference and decision making, those probability functions that she has not ruled out for such purposes.

One way of interpreting the classic distinction between decision making under risk and decision making under uncertainty (Luce and Raiffa 1957, 13) is to identify contexts in which the decision maker has precise probabilities over the relevant events with decision making under risk and contexts in which the decision maker's probabilities may be imprecise with decision making under uncertainty. By assuming agents should always have precise probabilities, decision making under uncertainty is reduced to decision making under risk by fiat. But Schervish and Seidenfeld's (1990) theorem 2 provides conditions under which learning or the acquisition of evidence will reduce uncertainty in the long run. Our proposition 1 establishes that certain evidence is not required to reduce uncertainty to risk eventually; such a reduction can be achieved with uncertain evidence as well.

Moreover, for an agent with a credal state of the same form as C, proposition 2 shows that, for any event $A \in \mathfrak{F}$, certain evidence is not required to stabilize on a precise opinion in the limit (e.g., $\mathbb{P}_\infty[A] = 0.8$) despite potentially extensive indeterminacy about the event initially (e.g., $\{\mathbb{P}[A] : \mathbb{P} \in C\} = [0.1, 0.9]$). Jeffrey conditioning can bring about such stability even though it may fail to lead to convergence to probabilistic certainty (0 or 1). This is an important distinction between the framework in which Schervish and Seidenfeld work and ours. They show that uncertainty can, in fact, be reduced to probabilistic certainty in the limit, while Jeffrey conditioning only guarantees that we can reduce uncertainty to risk given the appropriate assumptions.

9.3. Pooling and Learning and Learning and Learning. . . . A third upshot reinterprets the previous point in a social setting. Mathematical aggregation frameworks are general and precise settings in which to study ways of forming a *consensus* or group point of view from a set of potentially diverse points of view. Aggregation is an important topic in economics (e.g., allocation aggregation), political science (e.g., social choice theory), and statistics (e.g., opinion pooling), but it also finds application throughout decision theory and epistemology. Aggregation methods can be and often are interpreted as delivering a "consensus" position among sets of probabilities, beliefs, or preferences. Different concrete recipes for pooling probability judgments have been studied.[11] Some of the most extensively studied are ways of *averaging* probabilities to arrive at a group probability. For example, we could take linear or geometric averages of profiles of individual probabilities. We could also take such averages after weighting the probabilities in the profile differently. In general, pooling allows us to consider ways of arriving at a "consensus" apart from those of updating on a shared stream of evidence.

11. For a good survey, see Genest and Zidek (1986).

https://doi.org/10.1086/701954 Published online by Cambridge University Press

One assumption common throughout the pooling literature is that pooling produces a single "group" or "consensus" probability measure. Yet the standard frameworks have significant limitations. A number of results show that certain sets of desirable aggregation properties cannot be simultaneously satisfied on pain of triviality or inconsistency. Drawing on work on IP, Stewart and Ojea Quintana (2018b) motivate the use of IP in the context of pooling and generalize the canonical mathematical framework to allow for set-valued pooling functions. The group or consensus opinion is allowed to take the form of a *set* of probabilities. In this more general setting, a number of simple possibility results are established, showing that collections of desirable aggregation properties that are impossible to satisfy in the setting of precise opinion pooling can be jointly satisfied in the setting of opinion pooling with IP.

Stewart and Ojea Quintana also characterize a distinguished format of pooling with IP (2018b, proposition 6). That format is the function that returns the convex hull of any profile of probabilities to be aggregated.[12] The point of departure for concern with the convex hull—and the use of IP in the context of pooling more generally—is Levi's (1985) essay. There, Levi distinguishes between *consensus as shared agreement*, which is available at the outset of inquiry by retaining agreements and suspending judgment on other matters, and *consensus as the outcome of inquiry*, which emerges when evidence resolves initial disputes. About the role of the convex hull, Levi writes, "a potential resolution of the conflict between rival credal probability distributions is to be represented by a credal probability distribution which is the weighted average of the distributions in conflict. Hence, the set of all potential resolutions of such a conflict is to be represented by the convex hull (the set of all weighted averages) of the credal distributions initially in contention. My assumption was that this convex set of probability distributions represented the first kind of consensus I regard as important—consensus as shared agreements regarding probability judgment" (6). On this view, a pooling function that returns the convex hull can be regarded as delivering a consensus position at the outset of inquiry. This is a very different view of opinion pooling than typically presented.

One potential complaint about such a view of pooling is that, in many cases, it would identify rather weak consensus positions. According to Levi, this is as consensus at the outset of inquiry should be in order to avoid begging questions by suspending judgment between conflicting points of view. But that is not the end of his story. Weak points of view can be strengthened through inquiry. Levi writes, "if we adopt a consensus as shared agreement on credal probabilities before the acquisition of evidence, we may hope to obtain new data via experimentation and observation which will yield a con-

12. The framework in which convex IP pooling is characterized is general, however, subsuming any mapping from a profile of probabilities to a set of probabilities.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

sensus which resolves the original dispute via inquiry. In typical cases, *ample data will lead via Bayes's theorem and conditionalization to a reduction in the indeterminacy in the state of credal probability judgment. Consensus as the outcome of inquiry will be more determinate than the consensus as shared agreements adopted at the outset of inquiry*" (1985, 10, emphasis ours). One could see merging results as validating Levi's claim here. Under certain conditions, conditionalization does lead to a reduction in indeterminacy. And, happily enough, those conditions are met when forming the convex hull of some finite profile of mutually absolutely continuous probability functions to be aggregated into an initial consensus position. Furthermore, our proposition 1 shows that the ability to achieve the desired reduction of indeterminacy is not limited to Bayesian conditionalization. Under a few further assumptions, Jeffrey conditioning too reduces indeterminacy uniformly as evidence accumulates. At the outset of inquiry, we can "pool" opinions by identifying the group consensus with the convex hull of individual opinions. This weak initial position can be subsequently uniformly strengthened through inquiry to a consensus at the outcome of inquiry, as uncertainty is reduced to simple risk. Given our observations in this essay, Levi's picture of consensus appears to be robust against the choice of the updating method. Proponents of Jeffrey conditioning would reasonably balk at the account if the reduction of indeterminacy were a mere artifact of a learning rule that they think fails to capture so many learning experiences. We have shown that this is not the case.

It is also worth emphasizing again that any pooling function that outputs a subset of the convex hull of the profile presents a consensus position that is subject to all of the merging and convergence results for sets mentioned in this essay. If Levi is right that the convex hull represents "the set of all potential resolutions" of the conflict in probability judgments, and reasonable pooling functions take values in the power set of the set of such resolutions, then the merging and convergence results discussed here hold for the class of reasonable pooling functions. Merging results, we submit, constitute a partial response to complaints concerning the use of IP to identify a consensus at the outset of inquiry in the context of pooling.

Appendix

Proofs

Proof of Proposition 1. Let $\mathbb{P}^1, \ldots, \mathbb{P}^k$ be the extreme points that generate C. We show that for almost every $\omega \in \Omega$,

$$\sup_{\mathbb{P},\mathbb{Q}\in C} d(\mathbb{P}_n(\omega), \mathbb{Q}_n(\omega)) \to 0$$

https://doi.org/10.1086/701954 Published online by Cambridge University Press

as $n \to \infty$. Note that all of our "almost surely" qualifications hold for every $\mathbb{P} \in C$ by the mutual absolute continuity assumption.

Let $\mathbb{P}, \mathbb{Q} \in C$ be arbitrary. Following Huttegger (2015a, appendix), we note that for all $\omega \in \Omega$ the Jeffrey conditioning equation (6) can be written as

$$\mathbb{P}_n[\cdot](\omega) = \int \mathbb{P}[\cdot|\mathfrak{F}_n]dP_n(\omega). \tag{A1}$$

Using (A1) and our assumption that $P_n(\omega) = P_n^1(\omega) = Q_n(\omega)$ for almost every ω, we have

$$d(\mathbb{P}_n(\omega), \mathbb{Q}_n(\omega)) \leq \int d(\mathbb{P}_{\mathfrak{F}_n}, \mathbb{Q}_{\mathfrak{F}_n})dP_n^1(\omega) = \int d(\mathbb{P}_{\mathfrak{F}_n}, \mathbb{Q}_{\mathfrak{F}n}) \frac{dP_n^1(\omega)}{d\mathbb{P}^1} d\mathbb{P}^1 \tag{A2}$$

for almost all ω. The function $dP_n^1(\omega)/d\mathbb{P}^1$ is the Radon-Nikodym derivative of $P_n^1(\omega)$ with respect to $\mathbb{P}^1$ on the measurable space $(\Omega, \mathfrak{F}_n)$, which is guaranteed to exist for almost every ω because $P_n^1(\omega)$ is absolutely continuous with respect to $\mathbb{P}^1$ for almost every ω. Note that

$$d(\mathbb{P}_{\mathfrak{F}_n}, \mathbb{Q}_{\mathfrak{F}_n}) \leq \max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \text{ a.s.}$$

because $\mathbb{P}_{\mathfrak{F}_n}$ and $\mathbb{Q}_{\mathfrak{F}_n}$ are (a.s.) convex combinations of $\mathbb{P}^1_{\mathfrak{F}_n}, \ldots, \mathbb{P}^k_{\mathfrak{F}_n}$ (Schervish and Seidenfeld 1990). Therefore, for almost every ω,

$$\int d(\mathbb{P}_{\mathfrak{F}_n}, \mathbb{Q}_{\mathfrak{F}_n}) \frac{dP_n^1(\omega)}{d\mathbb{P}^1} d\mathbb{P}^1 \leq \int \max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \frac{dP_n^1(\omega)}{d\mathbb{P}^1} d\mathbb{P}^1.$$

From this inequality and (A2) we have

$$d(\mathbb{P}_n(\omega), \mathbb{Q}_n(\omega)) \leq \int \max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \frac{dP_n^1(\omega)}{d\mathbb{P}^1} d\mathbb{P}^1 \tag{A3}$$

for almost every ω. Since (A3) holds for all $\mathbb{P}, \mathbb{Q} \in C$, it holds upon taking a supremum over $\mathbb{P}, \mathbb{Q} \in C$. Hence, for almost all ω we have

$$\sup_{\mathbb{P},\mathbb{Q}\in C} d(\mathbb{P}_n(\omega), \mathbb{Q}_n(\omega)) \leq \int \max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \frac{dP_n^1(\omega)}{d\mathbb{P}^1} d\mathbb{P}^1. \tag{A4}$$

We conclude the proof with arguments similar to Huttegger's (2015a). We refer the reader to his appendix for more details. First, we note that (M′) implies that the sequence $\{dP_n^1(\omega)/d\mathbb{P}^1\}$ is a nonnegative martingale with respect to $\mathbb{P}^1$ for almost every ω, and hence converges a.s. to a finite limit for almost every ω. Theorem 2 implies that $\max_{i,j} d(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}) \to 0$ a.s. Therefore, for almost every ω,

$$\max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \frac{dP_n^1(\omega)}{d\mathbb{P}^1} \to 0$$

https://doi.org/10.1086/701954 Published online by Cambridge University Press

a.s. as $n \to \infty$. Using this fact we have

$$\lim_{n\to\infty} \int \max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \frac{dP^1_n(\omega)}{d\mathbb{P}^1} d\mathbb{P}^1 = \int \lim_{n\to\infty} \max_{i,j} d\left(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}\right) \frac{dP^1_n(\omega)}{d\mathbb{P}^1} dP^1 = 0 \tag{A5}$$

for almost every ω. The passage of the limit under the integral is justified by the fact that $dP^1_n(\omega)/d\mathbb{P}^1$ is uniformly integrable for almost every ω, which in turn implies that $\max_{i,j} d(\mathbb{P}^i_{\mathfrak{F}_n}, \mathbb{P}^j_{\mathfrak{F}_n}) dP^1_n(\omega)/d\mathbb{P}^1$ is uniformly integrable for almost every ω as it is dominated by $dP^1_n(\omega)/d\mathbb{P}^1$. That $dP^1_n(\omega)/d\mathbb{P}^1$ is uniformly integrable for almost every ω is equivalent to our assumption that $P^1_n(\omega)$ is uniformly absolutely continuous with respect to $\mathbb{P}^1$ for almost every ω. See Huttegger (2015a, proof of theorem 9.2) for more on this point.

Finally, (A4) and (A5) imply

$$\sup_{\mathbb{P},\mathbb{Q}\in C} d(\mathbb{P}_n(\omega), \mathbb{Q}_n(\omega)) \to 0$$

as $n \to \infty$ for almost every ω, as desired. QED

Jeffrey Conditioning Does Not Preserve Convexity

Proposition 3. Let C be as in proposition 1. The result of Jeffrey conditioning all elements of C on a common posterior need not be a convex set.

Proof. We sketch a counterexample. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, and consider the two probabilities in table A1.

TABLE A1. Priors

	ω_1	ω_2	ω_3	ω_4
$\mathbb{P}$	1/4	1/4	1/4	1/4
$\mathbb{Q}$	1/8	1/2	1/4	1/8

Set $C = \text{conv}\{\mathbb{P}, \mathbb{Q}\}$. Let $\mathfrak{E} = \{E_1, E_2\}$, with $E_1 = \{\omega_1, \omega_2\}$ and $E_2 = \{\omega_3, \omega_4\}$, be a partition of Ω. Jeffrey updating both $\mathbb{P}$ and $\mathbb{Q}$ using P, where $P(E_1) = 2/3$ and $P(E_2) = 1/3$, we obtain the posteriors in table A2.

TABLE A2. Posteriors

	ω_1	ω_2	ω_3	ω_4
$\mathbb{P}_1$	1/3	1/3	1/6	1/6
$\mathbb{Q}_1$	2/15	8/15	2/9	1/9

https://doi.org/10.1086/701954 Published online by Cambridge University Press

Let $C^P_{\mathfrak{E}}$ be the result of Jeffrey updating each element of C on the common posterior P on partition $\mathfrak{E}$. To establish the claim, it suffices to find some $\alpha \in [0, 1]$ such that $\alpha\mathbb{R}_1 + (1 - \alpha)\mathbb{R}'_1 \notin C^P_{\mathfrak{E}}$ for some $\mathbb{R}, \mathbb{R}' \in C$ (since it is clear that $\mathbb{R}, \mathbb{R}'_1 \in C^P_{\mathfrak{E}}$). Let $\alpha = 4/9$, and consider $\alpha\mathbb{P}_1 + (1 - \alpha)\mathbb{Q}_1$. Suppose for reductio that $C^P_{\mathfrak{E}}$ is convex. Then, $\alpha\mathbb{P}_1 + (1 - \alpha)\mathbb{Q}_1 \in C^P_{\mathfrak{E}}$. This implies that there is some $\beta \in [0, 1]$ and $\mathbb{R}^* \in C$ such that $\beta\mathbb{P} + (1 - \beta)\mathbb{Q} = \mathbb{R}^*$ and $\mathbb{R}^*_1 = \alpha\mathbb{P}_1 + (1 - \alpha)\mathbb{Q}_1$.

Clearly, $\alpha\mathbb{P}_1(\{\omega_3\}) + (1 - \alpha)\mathbb{Q}_1(\{\omega_3\}) = 16/81$. By the definition of Jeffrey conditioning and our assumptions, we have

$$\frac{16}{81} = \sum_{j=1,2} P(E_j)\mathbb{R}^*(\{\omega_3\}|E_j) = \left(\frac{2}{3}\right)\frac{\mathbb{R}^*(\{\omega_3\} \cap E_1)}{\mathbb{R}^*(E_1)} + \left(\frac{1}{3}\right)\frac{\mathbb{R}^*(\{\omega_3\} \cap E_2)}{\mathbb{R}^*(E_2)}.$$

Since $\{\omega_3\} \cap E_1 = \varnothing$, the left summand of the right-hand side is 0. By algebra and the definition of $\mathbb{R}^*$, we obtain

$$\frac{16}{81} = \left(\frac{1}{3}\right)\frac{\beta\mathbb{P}(\{\omega_3\}) + (1 - \beta)\mathbb{Q}(\{\omega_3\})}{\beta\mathbb{P}(E_2) + (1 - \beta)\mathbb{Q}(E_2)}.$$

Substituting values from table A1 and solving for β, we have $\beta = 3/8$. However, $\alpha\mathbb{P}_1(\{\omega_1\}) + (1 - \alpha)\mathbb{Q}_1(\{\omega_1\}) = 2/9$. This implies that $\mathbb{R}^*_1(\{\omega_1\}) = 2/9$. But for P and $\beta = 3/8$, this is not the case. Again using the definition of Jeffrey conditioning and our assumptions, it can be verified that $\mathbb{R}^*_1(\{\omega_1\}) = 22/111 < 2/9$. It follows that there do not exist a $\beta \in [0, 1]$ and a $\mathbb{R}^* \in C$ that meet our stated requirements above. QED

REFERENCES

Belot, G. 2013. "Bayesian Orgulity." *Philosophy of Science* 80 (4): 483–503.

Billingsley, P. 2008. *Probability and Measure*. New York: Wiley.

Blackwell, D., and L. E. Dubins. 1962. "Merging of Opinions with Increasing Information." *Annals of Mathematical Statistics* 33 (3): 882–86.

———. 1975. "On Existence and Non-existence of Proper, Regular, Conditional Distributions." *Annals of Probability* 3 (5): 741–52.

Cisewski, J., J. Kadane, M. Schervish, T. Seidenfeld, and R. Stern. 2018. "Standards for Modest Bayesian Credences." *Philosophy of Science* 85 (1): 53–78.

Diaconis, P., and S. L. Zabell. 1982. "Updating Subjective Probability." *Journal of the American Statistical Association* 77 (380): 822–30.

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

Elkin, L., and G. Wheeler. 2018. "Resolving Peer Disagreements through Imprecise Probabilities." *Noûs* 52 (2): 260–78.

Ellsberg, D. 1963. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 77 (2): 327–36.

Gaifman, H., and M. Snir. 1982. "Probabilities over Rich Languages, Testing and Randomness." *Journal of Symbolic Logic* 47 (3): 495–548.

Genest, C., and J. V. Zidek. 1986. "Combining Probability Distributions: A Critique and an Annotated Bibliography." *Statistical Science* 1 (1): 114–35.

Huttegger, S. M. 2013. "In Defense of Reflection." *Philosophy of Science* 80 (3): 413–33.

https://doi.org/10.1086/701954 Published online by Cambridge University Press

———. 2015a. "Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds." *Philosophy of Science* 82 (4): 587–601.
———. 2015b. "Merging of Opinions and Probability Kinematics." *Review of Symbolic Logic* 8 (4): 611–48.
Kalai, E., and E. Lehrer. 1994. "Weak and Strong Merging of Opinions." *Journal of Mathematical Economics* 23 (1): 73–86.
Kelly, K. T. 1996. *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.
Levi, I. 1974. "On Indeterminate Probabilities." *Journal of Philosophy* 71 (13): 391–418.
———. 1985. "Consensus as Shared Agreement and Outcome of Inquiry." *Synthese* 62 (1): 3–11.
———. 1987. "The Demons of Decision." *Monist* 70 (2): 193–211.
Luce, R. D., and H. Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey*. New York: Wiley.
Maher, P. 1992. "Diachronic Rationality." *Philosophy of Science* 59 (1): 120–41.
Savage, L. 1954/1972. *The Foundations of Statistics*. New York: Wiley.
Schervish, M., and T. Seidenfeld. 1990. "An Approach to Consensus and Certainty with Increasing Evidence." *Journal of Statistical Planning and Inference* 25 (3): 401–14.
Seidenfeld, T. 2001. "Remarks on the Theory of Conditional Probability: Some Issues of Finite versus Countable Additivity. In *Probability Theory: Philosophy, Recent History and Relations to Science*, ed. V. F. Hendricks. Dordrecht: Kluwer.
Škulj, D. 2006. "Jeffrey's Conditioning Rule in Neighbourhood Models." *International Journal of Approximate Reasoning* 42 (3): 192–211.
Skyrms, B. 1987. "Dynamic Coherence and Probability Kinematics." *Philosophy of Science* 54 (1): 1–20.
———. 1993. "A Mistake in Dynamic Coherence Arguments?" *Philosophy of Science* 60 (2): 320–28.
———. 1996. "The Structure of Radical Probabilism." *Erkenntnis* 45 (2–3): 285–97.
Stewart, R. T., and I. Ojea Quintana. 2018a. "Learning and Pooling, Pooling and Learning." *Erkenntnis* 83 (3): 369–89.
———. 2018b. "Probabilistic Opinion Pooling with Imprecise Probabilities." *Journal of Philosophical Logic* 47 (1): 17–45.
Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman & Hall.

https://doi.org/10.1086/701954 Published online by Cambridge University Press