Resolving Peer Disagreements Through Imprecise Probabilities

LEE ELKIN AND GREGORY WHEELER

Munich Center for Mathematical Philosophy, LMU Munich

Abstract

Two compelling principles, the Reasonable Range Principle and the Preservation of Irrelevant Evidence Principle, are necessary conditions that any response to peer disagreements ought to abide by. The Reasonable Range Principle maintains that a resolution to a peer disagreement should not fall outside the range of views expressed by the peers in their dispute, whereas the Preservation of Irrelevant Evidence (PIE) Principle maintains that a resolution strategy should be able to preserve unanimous judgments of evidential irrelevance among the peers. No standard Bayesian resolution strategy satisfies the PIE Principle, however, and we give a loss aversion argument in support of PIE and against Bayes. The theory of imprecise probability allows one to satisfy both principles, and we introduce the notion of a *set-based credal judgment* to frame and address a range of subtle issues that arise in peer disagreements.

1. Reasonable Range

You and a colleague hold different beliefs on the truth of the proposition that it will rain tomorrow in Riga. You think it is likely to rain. Your colleague believes otherwise. Neither you nor he can claim an epistemic advantage about the matter. You have the same evidence. The same level of expertise. The same powers of reasoning. You are *epistemic peers*. Upon learning that you have a disagreement with an epistemic peer, should you revise your beliefs? Should he? If so, how?

One response to an epistemic peer disagreement—or simply, *peer disagreement*—is to be conciliatory with your epistemic equals by adopting a new belief that assigns to each opinion in the disagreement equal weight (Elga 2007, p. 484) thereby splitting the difference (Christensen 2007, p. 203).[1] There are at least two versions of the equal-weight response, however, which engender different assumptions about the nature of the evidence a peer disagreement generates, and how that evidence should guide a peer to change her view.

According to most proponents of the equal-weight view a peer disagreement delivers to you evidence that either you or your peer is mistaken about the proposition in dispute. So, one version of the equal-weight view has it that the evidence from a

peer disagreement is undermining in character and therefore that your reaction to a peer disagreement ought to be the same as your reaction to receiving any other new but conflicting piece evidence: you ought to *suspend judgment* on the proposition until additional evidence is gathered (Feldman 2010).

If belief is interpreted categorically, suspending judgment on a disputed proposition amounts to neither believing it nor its negation. If instead belief is interpreted partially, and in particular is representable by a unique real-valued probability function, then suspension of judgment typically amounts to assigning a partial belief of $1/2$ to the proposition in question. Either way, the motivation for suspending judgment is the same. Since the evidence that one receives from a peer disagreement is taken to undermine rather than ameliorate one's current view, suspension-of-judgment versions of the equal-weight view are guided by the notion that one ought to respond to a peer disagreement by increasing one's uncertainty about the proposition in dispute.

Another version of the equal-weight view counsels against suspending judgment. On this version a peer disagreement supplies you with a range of informed opinions, including your own, so you ought to exploit this information to improve upon your current judgment. Here the evidence from a peer disagreement is taken to be ameliorative in character, so one ought to respond by taking the equally-weighted *average* of the set of peer judgments as one's new partial belief (Douven 2010).[2]

One advantage opinion pooling strategies have over a naïve suspension of judgment is that pooling strategies in general, and equally-weighted averaging in particular, yield a new partial belief that is guaranteed to fall within the *reasonable range* of informed opinions.

> **Reasonable Range Principle:** For any group of peers, $\mathbb{P}$, whose partial beliefs in a proposition A range from x, the lowest confidence in the truth of A expressed by a member $\mathbb{P}$, to y, the highest confidence in the truth of A expressed by a member of $\mathbb{P}$, a new belief is said to be within the *reasonable range* for members of $\mathbb{P}$ if and only if its value is within the closed interval $[x, y]$.

To motivate why the Reasonable Range Principle is reasonable and a policy of naïve suspension of judgment is not, imagine that your degree of belief in rain tomorrow in Riga is $8/10$ but your epistemic peer's is $9/10$. Upon learning of this disagreement it would be foolish to advise either you or your peer to naïvely suspend judgment by adopting a partial belief of $1/2$ that it will rain tomorrow. After all, you both agree that it is more likely to rain in Riga than it is for a fairly tossed coin to land heads, and no strategy to resolve a disagreement among peers should mandate that each ought to suspend judgment on a proposition they both believe is overwhelmingly more likely to be true than false. Whatever uncertainty this peer disagreement may introduce, it should not wipe out this point of agreement.

One may nevertheless worry that some disagreements warrant adopting a position outside the range prescribed by the Reasonable Range Principle. Discovering you are party to a disagreement introduces you to variance where there was comparatively little or none before, and sometimes the reasonable response to a channel of

information that increases your variance is to fault the channel rather than submit to constraints imposed by the information it delivers.[3] For example, Christensen (2009, p. 759) describes an agent who is confident but not dogmatic that a treatment dosage to a patient is correct (0.97) yet responds to a colleague's strong but slightly less confident belief in the same (0.96) by *boosting* his confidence that the dosage is right. But this is an unreasonable response to an epistemic peer. For while the agent in Christensen's example expresses low confidence that the administered treatment is the incorrect dosage, his colleague is slightly less confident that this error is avoided. Yet if the confidence-boost response were right, the agent would be licensed to infer from his colleague's judgment that the prospect of administering the incorrect dosage is now *lower* than he originally believed, which is absurd unless he views his colleague's judgment to be biased away from the truth in a way that his own judgment is not. That's no way to treat a peer.

Responding to a disagreement by adopting a judgment that falls outside the range of group opinion is reasonable only if your colleagues are not your epistemic peers. For if every peer's partial belief that A is between x and y, where $[x, y]$ is the smallest span covering every peer's judgment, then violating the Reasonable Range Principle entails adopting a belief whose value is outside the range considered reasonable by one's peers. A response to a peer disagreement which did not satisfy the Reasonable Range Principle would in effect either deny that the disagreement is among epistemic peers or license one to deliberately move away, without reason, from the considered opinions of her peers.

That said, equally-weighted averaging is not the only response to a peer disagreement that satisfies the Reasonable Range Principle. This is fortunate since there are cases where it is unreasonable to resolve a disagreement among peers by taking some or another *non-extreme* weighted average of peer opinions.[4] If you are party to a peer disagreement in which nine out of ten agree yet one outlier does not, the reasonable response may be for the outlier to fall in line with the majority rather than for the majority to move partway to meet the outlier. Peerage does not confer infallibility, after all. Sometimes what a peer learns from a disagreement with his equals is that he is in the wrong. We will return later to discuss 'higher-order' evidence that a group disagreement can produce. For the moment, we only wish to point out that allowing a single peer to change his view to join a steadfast majority is a case where the Reasonable Range Principle is satisfied but non-extreme weighted averaging is not. In fact, any 'permissive' response to peer disagreement which allows a party to a disagreement to stick to her original judgment will trivially satisfy the Reasonable Range Principle.[5]

Even though the Reasonable Range Principle is satisfied by a variety of competing peer disagreement strategies—including Savage's Minimax, calibrated maximum entropy, Maximax, and Levi's E-admissibility—classical Bayesian methods that satisfy the Reasonable Range Principle nevertheless appear to rule out an important insight from the suspension-of-judgment view, namely that at least some peer disagreements deliver greater uncertainty to each member of the group. It is wishful thinking to suppose evidence from every peer disagreement to be ameliorative in character: sometimes the correct response to a peer disagreement is to be more

uncertain about the proposition in dispute. But if it is true, as we will return to argue, that some peer disagreements warrant a response that increases one's uncertainty, how can a peer's newfound uncertainty from a peer disagreement be reconciled with the Reasonable Range Principle? That is one of the questions we address in this paper.

Another set of questions we will address concerns a problem Bayesian views have preserving some shared points of agreement among peers. It is this latter issue we turn to next.

2. Preservation of Irrelevant Evidence

Discussions of peer disagreement typically focus exclusively on the special case of two peers disputing a single proposition,[6] thus ignoring other forms a peer disagreement may take and the different responses each form may warrant. For instance, a single outlier capitulating to his nine peers illustrates how the distribution of group judgments may yield evidence warranting some members of the group to respond differently than others. One motivation for restricting attention to two-peer disagreements, however, is precisely to set aside disagreements that are easily defused by 'swamping' higher-order evidence (Kelly 2010). Since higher-order evidence is not always available, the restriction to two peers helps to bring the problem of peer disagreement into sharper focus.[7]

The same however cannot be said for restricting attention to a single proposition. Any proposal for resolving a peer disagreement involving one proposition should be able to handle a disagreement involving two. Yet, two peers disagreeing over two propositions puts the kibosh on non-extreme weighted averaging strategies. To see why, consider the following modification of our weather forecasting example.

> **Heads and Rain Example:** *Suppose that Meteorologist One and Meteorologist Two share the same data provided by the* European Center for Medium Range Weather Forecasting *and they each use this data to forecast rain in Riga for the following day* (R). *Meteorologist One's partial belief in R is 0.40 and Meteorologist Two's partial belief in R is 0.55. Included in their shared knowledge is information about a biased coin to be flipped today and the two meteorologists disagree about that outcome, too. One's credence in the coin landing heads today (H) is 0.2, while Two's credence in H is 0.6. Even so, both agree that rain tomorrow in Riga and the coin landing heads today are stochastically independent. So, while the meteorologists disagree on rain tomorrow and they disagree on the coin landing heads today, both agree that there is no value in knowing the outcome of the coin flip to forecasting rain tomorrow in Riga.*

Both Meteorologist One and Meteorologist Two believe that rain in Riga tomorrow and the coin landing heads today are *stochastically independent*: that is, both $p_1(R \land H) = p_1(R)p_1(H)$ and $p_2(R \land H) = p_2(R)p_2(H)$, where p_1 and p_2 represent the partial beliefs of Meteorologist One and Meteorologist Two, respectively. So, however One and Two decide to resolve their disagreements about today's coin flip

Table 1. Forecasters p_1 and p_2 and their equally-weighted average p^*

	$p_1(\cdot \wedge \cdot)$	$p_2(\cdot \wedge \cdot)$	$p^*(\cdot \wedge \cdot)$	$p^*(\cdot)p^*(\cdot)$
H R	0.08	0.33	0.205	0.19
H R^c	0.12	0.27	0.195	0.21
H^c R	0.32	0.22	0.27	0.285
H^c R^c	0.48	0.18	0.33	0.315

and tomorrow's weather, their resolution should preserve the judgment that heads today yields *irrelevant evidence* for forecasting rain tomorrow.

> **Preservation of Irrelevant Evidence (PIE) Principle:** If every member of a group of peers, $\mathbb{P}$, believes that her partial belief in the truth of proposition A should remain unchanged whether or not another proposition B is true, and no member of the group changes her mind about the irrelevance of B to A after the disagreement becomes common knowledge to the group, then the resolved peer disagreement should preserve the judgment that B is irrelevant evidence to A.

The problem is that *any* non-extreme weighted average of p_1 and p_2 that One and Two might propose to resolve their disagreement will violate the PIE Principle. Without loss of generality, consider the specific case of p^* in Table 1, which is the equally weighted average of p_1 and p_2, i.e., $p^* = \frac{1}{2}p_1 + \frac{1}{2}p_2$. The 'middle-ground' determined by p^* fails to preserve independence between the coin toss today and the weather tomorrow,[8] so resolving One and Two's disagreements by p^* would not satisfy the PIE Principle.

Although we find the PIE Principle intuitively compelling, not everyone agrees. Lehrer and Wagner (1983), for instance, have argued that violations of the PIE Principle are of "negligible epistemic significance." Even critics of weighted averaging schemes, such as Jon Williamson (2015), would ague that the PIE Principle should not constrain rational belief. But flouting the PIE Principle is not merely unintuitive; it is irrational.

To see why violating the PIE Principle is irrational, suppose One and Two reconcile their disagreement by p^* yet persist in believing that heads today is irrelevant to the event of rain tomorrow. A clever gambler may then compel them to accept a contract consisting of the following two bets.[9] The first bet consists of the gambler buying from the peers a ticket, T1, for €20.50, that pays the gambler €100 if the coin lands heads today and it rains in Riga tomorrow, and pays him nothing otherwise. The second bet consists of the gambler buying from the peers a second ticket, T2, for €60.75, that pays the gambler €225 if the coin lands tails today and it rains in Riga tomorrow, and pays him nothing otherwise. The payoffs to the gambler for each possible outcome are given in Table 2.

According to p^*, both T1 and T2 are considered fair by the reconciled peers.[10] Yet, since both the gambler and the peers agree that the coin flip today yields

Table 2. Gambler's payoffs

	Ticket 1	Ticket 2	Net
$H \wedge R$	€79.50	− €60.75	€18.75
$H \wedge R^c$	− €20.50	− €60.75	− €81.25
$H^c \wedge R$	− €20.50	€164.25	€143.75
$H^c \wedge R^c$	− €20.50	− €60.75	− €81.25

irrelevant evidence to the weather tomorrow, the gambler may opt to determine his payoff according to the product of the pooled marginal probabilities for each state by swapping the values in the third column of Table 1 for the values in the fourth. But now the gambler's expected gain according to the swapped values is positive.[11] So, no matter which of the four states obtains, the peers' expected payoff is now negative. Finally, suppose the gambler compels the peers to accept two more called-off bets in the same spirit as the first pair, only now the bets are arranged for the peers to judge as fair a pair of bets under the product of the pooled marginal distributions that incurs an expected loss on the pooled joint distribution. For instance, suppose the gambler sells to the peers a ticket T3 for €13.30 that pays them €70 on heads and rain but zero otherwise, and he sells to them another ticket T4 for €32.20 that pays them €120 on tails and rain but zero otherwise. Then, with this contract of four bets, T1–T4, the peers are booked in an expected sure loss whichever way they decide to resolve their bets with the gambler.

One way around this argument is to double-down on weighted averaging by adopting the new betting odds given by p^* as rational and deny that the coin and weather should remain independent in the reconciled judgment. However, this response would enjoin the peers to place some value in the information provided by today's coin flip to further their epistemic goal of forecasting tomorrow's weather. So, according to this line, it would be rational for the peers to pay a fee, even if only a fraction of a cent, to learn the outcome of today's coin flip in order to better forecast tomorrow's weather. This is clearly absurd. While this move forecloses the possibility of suffering a sure loss (in expectation), it opens another for snake oil salesmen to sell to the peers epistemically useless information.

The upshot of this argument is a dilemma for conciliatory Bayesians. On the one hand the measure p^*, which is the obvious function for the Bayesian version of the equal-weight view, cannot preserve independence. Thus, the Bayesian equal-weight view cannot accommodate the PIE Principle. This argument applies to any conciliatory Bayesian who adopts a non-extreme weighted averaging of probabilities, and it extends to other conciliatory methods that fail to preserve independence.[12] On the other hand a conciliatory Bayesian who rejects the PIE Principle is committed to the view that a shared judgment of irrelevance among peers cannot, and should not, be preserved by any resolution strategy. Thus, the Bayesian without PIE becomes a mark for swindlers and soothsayers.

One way to escape the Bayesian's dilemma is simply to permit extreme weighted averaging. But this amounts to conciliation by ultimatum: you can hold any opinion you like so long as it is mine. This response is hardly a conciliatory strategy. Without a principled reason for picking one peer's judgment over another, there is little to recommend the ultimatum strategy for resolving a disagreement among peers. Another response is simply to leave the set of peer judgments unchanged. Each peer in the set would satisfy the PIE Principle by digging in her heels and rejecting any change to her view. To be clear, we do not think there is a compelling argument for the view that it is always rational to respond to a peer disagreement by remaining steadfast. Without appealing to higher-order evidence, it is difficult to conceive of adequate grounds to warrant picking one view over others or remaining steadfast, and it is doubtful that such higher-order evidence is always available.

Before continuing to consider an argument against non-conciliatory responses to peer disagreements, which we do in Section 4, a natural question to ask is whether there is some other option for reconciling the PIE Principle with the demands of conciliation. The short answer is, Yes: it is straight-forward to formulate concilia-tory responses to peer disagreements with imprecise probability theory that satisfy both the Reasonable Range Principle and the PIE Principle. The first conclusion to draw from our approach, which we introduce in the next section, is that one should question approaches that mandate a single determinate credal probability long before calling into question conciliatory responses that satisfy the PIE Principle.

3. Set-based Credal Judgments

Stripped of bells and whistles, a set-based credal judgment is a straight-forward extension of numerically determinate degrees of belief pioneered by Frank Ramsey (1926) and Bruno de Finetti (1937, 1974). Mathematically, degrees of belief (or partial beliefs or credences) are represented canonically by a finitely additive prob-ability function, p, that assigns to events of an algebra $\mathscr{A}$ over a (finite) set of states Ω a real number between 0 and 1. A set-based credal judgment, to be explained in this section, is represented in terms of a (non-empty) set $\mathbb{P}$ of probability functions, each defined with respect to the same structure, $(\Omega, \mathscr{A})$. For the moment, one may think of $\mathbb{P}$ as a set of Bayes agents, or set of peers, each with her own view about a set of propositions $\mathscr{A}$.

In our heads and rain example, $\mathbb{P} = \{p_1, p_2\}$ represents the judgments that Meteorologist One and Meteorologist Two have regarding the state of affairs of the coin landing heads today and the event of rain in Riga tomorrow. The reason-able range of opinions on whether it will rain tomorrow in Riga is from 0.4 to 0.55, and likewise the reasonable range is 0.2 to 0.6 for today's coin flip landing heads. Generally, for each event E in $\mathscr{A}$, there is some probability p in $\mathbb{P}$ whose value is the smallest of any in $\mathbb{P}$, which is the *lower probability* of E, and some p in $\mathbb{P}$ whose value is the largest of any in $\mathbb{P}$, which is the *upper probability* of E.

We adopt a common abuse of notation by identifying the proposition E with the indicator of the event E occurring, writing for example $p_1(E) = 0.4$ instead

of $p_1(\mathbf{1}_E(\omega) = 1) = 0.4$ and $p_1(\neg E) = 0.6$ instead of $p_1(\mathbf{1}_E(\omega) = 0) = 0.6$, where $\omega \in \Omega$ and $E \subseteq \mathscr{A}$ such that

$$\mathbf{1}_E(\omega) = \begin{cases} 1 & \text{if } \omega \in E, \\ 0 & \text{if } \omega \notin E. \end{cases}$$

Call the quadruple $(\Omega, \mathscr{A}, \mathbb{P}, \underline{P})$ a *lower probability space*, where Ω is a set of states, $\mathscr{A}$ is an algebra over Ω, $\mathbb{P}$ is a nonempty set of probability functions on $\mathscr{A}$, and $\underline{P}$ and $\overline{P}$ are functionals on $\mathscr{A}$ such that for each event E in $\mathscr{A}$:

(**Lower probability**) $\underline{P}(E) = \inf \{\, p(E) : p \in \mathbb{P} \,\}$,

(**Upper probability**) $\overline{P}(E) = \sup \{\, p(E) : p \in \mathbb{P} \,\}$.

Lower probability and upper probability satisfy a conjugacy relation, $\overline{P}(E) = 1 - \underline{P}(\neg E)$, which means that we only need to specify one of the two functionals. By convention, the lower probability $\underline{P}$ is usually specified.

If F is in $\mathscr{A}$ and $\underline{P}(F) > 0$, then conditional lower probabilities and conditional upper probabilities are defined as

(**Conditional lower probability**) $\underline{P}(E \mid F) = \inf \{\, p(E \mid F) : p \in \mathbb{P} \,\}$,

(**Conditional upper probability**) $\overline{P}(E \mid F) = \sup \{\, p(E \mid F) : p \in \mathbb{P} \,\}$.

If F is the sure event Ω, conditional lower probability and conditional upper probability reduce to unconditional lower probability and unconditional upper probability, respectively. For the remainder, assume that all lower and upper probabilities are defined with respect to the same lower probability space. We also omit reference to the underlying space $(\Omega, \mathscr{A})$ when the context is clear.

When the lower and upper probabilities are the same for all events in the algebra, we say that the peers (so represented) are in *full agreement*. When a set of peers are in full agreement, the set $\mathbb{P}$ is a singleton set consisting of a unique probability function realizing the upper and lower probabilities for every event:

(**Full agreement**) If $\underline{P} = \overline{P}$, then $\{p\} = \mathbb{P}$ and $p = \underline{P} = \overline{P}$.

A peer disagreement therefore occurs just when there is at least one event for which the upper and lower probabilities are not equal.

(**Peer disagreement**) Let $\underline{P}$ and $\overline{P}$ be defined with respect to lower probability space $(\Omega, \mathscr{A}, \mathbb{P}, \underline{P})$. A *peer disagreement* among $\mathbb{P}$ occurs if and only if there is some $E \in \mathscr{A}$ such that $\underline{P}(E) \neq \overline{P}(E)$.

Lower and upper probabilities are an old idea, dating back at least to Bernoulli (1713) and Boole (1854), and developed further by Koopman (1940) and Halmos (1950). After World War II, it was observed that the language of events and *lower probabilities* is more limited in expressive capacity than the language of random variables and (lower) expectations or *lower previsions* (Smith 1961; Williams 1975;

Table 3. Reasonable Ranges and Lost Independence

		H	R	$R \mid H$	$R \wedge H$	$R \wedge H$
(a)	p_1	0.2	0.4	0.4	.08	$\mathbb{P}_a[0.08, 0.33]$
	p_2	0.6	0.55	0.55	.33	
(b)	p_3	0.2	0.4	0.55	.11	$\mathbb{P}_b[0.11, 0.24]$
	p_4	0.6	0.55	0.4	.24	

Walley 1991), an observation that has several far-ranging consequences. Neverthe-less, we set those developments to one side in this paper and restrict ourselves to a very simple lower probability model.

Lower probability spaces provide a general framework within which to represent and evaluate a variety of responses to peer disagreements. Every probabilistic account for peer disagreement that we are aware of that satisfies the Reasonable Range Principle can be represented and compared within this setting. As we indicated above, a lower probability space whose basis is a singleton set of one probability function is equivalent to a standard, numerically determinate probability model. In our setting this model is the model of full agreement, and the Bayesian view of reconciling peer disagreement is simply one of specifying the method whereby a new model of full agreement is selected.

Although we are not the first to advocate using imprecise probability to model opinion pooling in general (Walley 1981) and group disagreement in particular (Levi 1990), our approach pays particular attention to the structural properties of the underlying set of probabilities that form the basis for lower and upper probability assessments. As we will argue, this basis for upper and lower probability judgments plays a crucial role in modeling group opinions. Unlike a classical Bayesian model, where all of the epistemically relevant information about an agent's cognitive commitments is allegedly captured by a single, numerically precise probability function, lower and upper probability functions alone do not capture all epistemically relevant information about an agent's cognitive commitments. Unlike the approaches of Levi (1980) or Walley (1991), who are committed to closed convex sets of probabilities either as a consequence of rationality principles (Levi) or for mathematical expediency (Walley), our position is that convex bases ought to be permitted but not mandated.

Intuitively, a *set-based credal judgment* toward a proposition E induces a lower and upper probability of E. To see why the reasonable range determined by lower and upper probabilities fails to capture all the information relevant to a peer disagreement, consider again the heads and rain example.

The judgments of Meteorologist One and Meteorologist Two are displayed in the top two rows of Table 3 labeled (a). The bottom two rows, labeled (b), describe a different pair of Meteorologists, Three and Four. The last column of Table 3 gives the reasonable range of One and Two's judgments on the joint event of heads and

rain, $\mathbb{P}_a(R \wedge H) = [0.08, 0.33]$, followed by Three and Four's reasonable range of the same joint event, $\mathbb{P}_b(R \wedge H) = [0.11, 0.24]$. One and Three hold identical views on heads today and on rain tomorrow, which are different than the shared view of Two and Four on those same two events. However, group (a) differs from group (b) in the conditional judgments they endorse. For group (a), the observation of heads today is irrelevant information to forecasting rain tomorrow. For group (b), the outcome of heads today does provide relevant information to forecasting rain in Riga tomorrow but Three and Four disagree with one another over how: Three believes that heads and rain are positively correlated, whereas Four believes they are negatively correlated. Despite this difference between group (a) and group (b), all four have the same reasonable range for the conditional judgment of rain given heads: $\mathbb{P}_{a \cup b}(R \mid H) = [0.4, 0.55]$.[13]

Although the reasonable ranges for the separate events of heads and rain and the reasonable range of the conditional judgment of rain given heads cannot distinguish group (a) from group (b), the reasonable ranges for the joint event of both rain and heads do reveal a difference between the two groups: that is, $\mathbb{P}_a(R \wedge H) \neq \mathbb{P}_b(R \wedge H)$. So far, so good. However, if we were to pool (a) and (b) into a single group, the reasonable range for Three and Four on heads would be properly included in the reasonable range of One and Two's judgment. We then would be unable to distinguish between the merged group and the original pair by the reasonable range of opinions alone. This point generalizes. Say that R is *irrelevant* to H just in case both $\underline{P}(R \mid H) = \underline{P}(R \mid \neg H) = \underline{P}(R)$ and $\overline{P}(R \mid H) = \overline{P}(R \mid \neg H) = \overline{P}(R)$, where R and H are each non-zero probability events. Say H and R are *epistemically independent* when both H is irrelevant to R and R is irrelevant to H. In general, if H is epistemically independent to R under $\underline{P}$, it does not follow that H and R are stochastically independent under every p in $\mathbb{P}$.[14]

Fortunately, the converse holds. That is, if R and H are stochastically independent under every p in $\mathbb{P}$, then R and H are epistemically independent under $\underline{P}$. In the parlance of imprecise probability theory, $\underline{P}$ defined in this manner is an *independent lower envelope* (Walley 1991, p. 446). Notice that in our original example, where $\mathbb{P}$ consists of just p_1 and p_2, $\underline{P}$ is an independent lower envelope, but adding either p_3 or p_4 to $\mathbb{P}_a$ destroys this property. While the reasonable ranges for $\underline{P}$ on $\mathbb{P}_{a \cup b}$ are the same as the reasonable ranges for $\underline{P}$ on $\mathbb{P}_a$, not every p in $\mathbb{P}_{a \cup b}$ judges the two events independent.

So far we have merely introduced some notation to make precise a bit of common sense. If One and Two agree that heads today and rain tomorrow are irrelevant to one another, adding someone else to the group who believes otherwise would break that consensus. But this notation allows us to specify a variety of commitments that a group of peers may have, and to work out the sometimes subtle consequences that follow from them.[15]

For instance, return to the original heads and rain example. One and Two each judge heads and rain to be independent and their shared judgment of irrelevance becomes common knowledge to them upon learning of their disagreement. That is to say, since every p in $\mathbb{P}_a$—hereafter we return to writing $\mathbb{P}$ instead of $\mathbb{P}_a$—renders R independent of H, the basis set $\mathbb{P}$ satisfies the conditions of an independent

lower envelope. So, the peers' individual *ex ante* judgments of epistemic irrelevance between H and R in $\mathbb{P}$ ensure that their (shared) *ex post* credal judgments determined by $\underline{P}$ and $\overline{P}$ render H epistemically irrelevant to R and R epistemically irrelevant to H.

By contrast, if we replaced the two-element set $\mathbb{P}$ by its convex hull, co $\mathbb{P}$,[16] then $\underline{P}$ based on co $\mathbb{P}$ would not be an independent lower envelope, even though the distributions in co $\mathbb{P}$ which realize $\underline{P}$ and $\overline{P}$ satisfy epistemic independence. Here too we are merely redescribing a familiar point in different terms, for the difference between the original set $\mathbb{P}$ and its convex hull co $\mathbb{P}$ is precisely the open set of all possible non-extreme weighted averages of p_1 and p_2. Further, as a terminological aside, but one that may help connect together some of the disparate communities working on imprecise probability, the convex hull of $\mathbb{P}$ corresponds to Walley's *natural extension* of $\mathbb{P}$ (1991), Levi's *credal set* (1980), and Joyce's *credal committee* (2010). From one point of view, the natural extension is the most generic technique for constructing credal judgments and conditional credal judgments because it ignores various structural judgments that may be present in the original set $\mathbb{P}$. Walley discusses different extensions that incorporate different structural judgments, yielding what Haenni et al. call different *parameterizations* of a set of probabilities (2011). The independent lower envelope is one of them. There are others (Augustin, Coolen, de Cooman, and Troffaes 2014).

Returning to our discussion of set-based credal judgments, we are now in a position to say what it means for a credal judgment determined by $\underline{P}$ and $\overline{P}$ to be based on a set $\mathbb{P}$.

> (**Set-based credal judgment**) Given a lower probability space $(\mathbf{\Omega}, \mathscr{A}, \mathbb{P}, \underline{P})$, a set-based credal judgment for an event E in $\mathscr{A}$ is determined by $\mathbb{P}$ and the pair $\underline{P}(E)$ and $\overline{P}(E)$. We say that $\mathbb{P}$ is the basis for the credal judgment E determined by $\underline{P}$ and $\overline{P}$. Similar remarks extend to set-based conditional credal judgments.

The point of set-based credal judgments is this. When assessing a credal judgment determined by $\underline{P}$ and $\overline{P}$ in the manner we have introduced here, one must bear in mind the underlying lower probability space $(\mathbf{\Omega}, \mathscr{A}, \mathbb{P}, \underline{P})$, including the structure of $\mathbb{P}$.[17] Fortunately, peer disagreements as we define them in this paper supply the information necessary to specify each component of a lower probability space, including the structure of $\mathbb{P}$. And these features allow one to work out subtle differences among a variety of judgments.

For example, suppose a group of peers disagree over judgments of evidential relevance. This case arose when we added Meteorologists Three and Four to the original pair of peers. But there are also cases where an unanimous *ex ante* judgment of independence should not be preserved in the group's *ex post* judgments. In other words, there are cases where a group of peers is initially in agreement that two events are stochastically independent but learning they are in disagreement over some probability judgment destroys this consensus and warrants the peers to reject their initial judgments of independence and to affirm that one event is relevant to the other. This possibility is the reason why the PIE Principle includes the

provision that no member of the group change her mind once the disagreement becomes common knowledge.

To see how common knowledge of a disagreement can undermine a prior judgment of irrelevance, imagine two urns that both contain the same number of red and white balls. Specifically, suppose there are 99 balls of one color and a single ball of the other color in both urns, and suppose this is common knowledge to two peers named Five and Six. Peer Five believes that both urns contain 99 red balls and 1 white ball, whereas peer Six believes that both urns contain 99 white and 1 red. Both Five and Six believe, falsely, that they are in agreement about the composition of the two urns; neither considers it *ex ante* to be a serious possibility that they may disagree. So, each peer's *ex ante* belief about the urns is that a randomly drawn ball from the first urn is evidentially irrelevant for estimating the probability of drawing a red ball from the second urn. Now suppose the peers discover their disagreement with one another. Then, each peer will believe *ex post* that a randomly drawn ball from the first urn is highly relevant for estimating the probability of drawing a red ball from the second urn. In this case their *ex ante* judgments of independence should not be preserved in their *ex post* judgments.

The difference between the original heads and rain example and the two urns example is that in the first example no member of the group changes her mind about any structural judgment of irrelevance upon discovering their disagreement but in the second example everyone changes her mind about relevance upon discovering their disagreement. Notice, however, that the bases for the heads and rain example and for the two urns example both generate independent lower envelopes. What differentiates the original heads and rain example from the two urns example is that One and Two in the original example maintain the judgment that the marginal probabilities of heads and rain are independent, whereas this condition is not applicable to the two urns example and thus not binding on Five and Six. In the parlance of imprecise probability theory, these two examples illustrate the difference between *strong independence* and independent lower envelopes (Miranda and de Cooman 2014): an independent lower envelope satisfies strong independence if the marginal distributions are stochastically independent. So, while the representations of the original heads and rain example and the two urns example both satisfy the conditions for an independent lower envelope, only the representation of the heads and rain example satisfies the additional condition necessary for strong independence.

4. Why Set-based Credal Judgments?

A set-based credal judgment is one where an agent's credal commitment toward a proposition induces a lower and upper probability representation of that commitment to the proposition. In the last section we cautioned against the mistake of simply identifying an agent's credal commitments with the interval induced by $\underline{P}$ and $\overline{P}$ for an event. One must also attend to the parameterization of $\underline{P}$, which will be reflected both in the original topological structure of $\mathbb{P}$ and by judgments made about properties of an extension that should or should not be preserved in light of a disagreement. Although the choice of extension for $\underline{P}$ is foreign to a traditional

Bayesian, this degree of freedom is merely a byproduct of the increased expressive capacity of imprecise probability.

That said, set-based credal judgments typically do yield something resembling an interval of credal opinion. The lower probability and upper probability for rain tomorrow in Riga, R, induced by p_1 and p_2 from our original example, yields an interval constraint (of some kind) pictured like so:
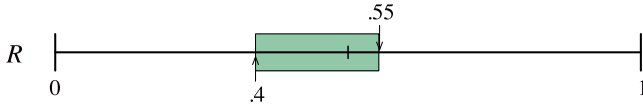
We are calling the interval between 0.4 and 0.55 the reasonable range of opinion on R, but others have appealed to the idea of a credal committee (Joyce 2010; Bradley 2014) or mental committee (Moss 2015), which are simply alternative names for a credal set (Levi 1980). The very rough idea is that the span between 0.4 and 0.55 captures some important features of indeterminacy in opinion, or imprecision in elicitation, that cannot be expressed by a determinate probability. In the peer disagreement problem, an indeterminate judgment for some proposition is imposed on each peer after she receives news of equally credible estimates that nevertheless are at variance with her own initial judgment. For us, unlike Levi and his followers, we do not mandate convexity.

Since the peer disagreement problem is traditionally assumed to involve a group of Bayes agents, each peer's original credal judgment is a precise partial belief.[18] What this means, in the traditional Ramsey-de Finetti conception of degrees of belief, is that each peer has a fair price for the proposition in question. In other words, what it means for Forecaster One to have a degree of belief of 0.4 in the proposition expressing that it will rain tomorrow in Riga is that he is indifferent to engaging in two types of transactions. The first hypothetical transaction calls on him to *buy* a contract for €0.40 that pays him €1 if it rains in Riga and nothing otherwise; the second hypothetical transaction calls on him to *sell* a contract for the same price. To unpack this further, when an agent agrees to buy such a contract, what he agrees to do is surrender a sure reward of 40 cents to acquire the uncertain reward of 1 Euro on the condition that it rains in Riga tomorrow. Similarly, when an agent agrees to sell such a contract, what he agrees to do is surrender a contract that gives him the uncertain reward of 1 Euro on the condition that it rains to acquire the sure reward of 40 cents. According to this tradition, an agent's credal judgment can be identified with his commitment to a system of fair prices for buying and selling any finite number of contracts. The agent's commitment is rational if and only if the resolution of the bets behind such contracts does not incur a sure loss for him if and only if the prices he commits to satisfy the axioms of finitely additive probability.

We rehearse this canonical account in order to point out something that parties to a peer disagreement learn. By announcing a fair price of 0.4 that R, Forecaster One announces that he is unwilling to pay more than 40 cents for a contract that returns to him 1 Euro in the event of R, and Forecaster Two learns from this signal that she, Forecaster Two, may have overpriced R. Think about how One and Two

would respond to unit gambles on R offered to them for less than 40 cents: both would snap them up as bargains. So, the span from 0 to 0.4 may be viewed as the range of agreement on buying prices for unit bets on R: each peer would respond to offers within this range in exactly the same way, since each judges the expected value of $(R - \alpha)$ to be greater than 0 for buying prices α of 40 cents or less. The two peers differ, however, in how they respond to offers for unit gambles on R that are priced between 40 and 55 cents. Forecaster One would not buy a unit gamble on R in this price range, whereas Forecaster Two would. Because they are epistemic peers, Two receives this news from One as a signal that she may be disposed to pay too high a price for this bet on R. Therefore, Two's buying price for a unit gamble on R should change to agree with One's. This is simply what it means for Forecaster Two to change her original buying price expressed by her initial probability of 0.55 for R to the lower probability of 0.4 for R.

Roles are reversed when we turn to the selling price for R. Here Forecaster Two will not surrender a gamble on R that pays her 1 Euro on the event of Rain in Riga to acquire in exchange a sure reward of any amount less than 55 cents, whereas Forecaster One is willing to sell a unit gamble on R for as low as 40 cents. Forecaster One therefore is committed to unloading contracts on R for a price that Forecaster Two would never agree to match. Now the standpoints of the two peers are reversed. Whereas both One and Two would agree to sell a gamble returning the uncertain reward of 1 Euro for a sure reward of 55 cents or more, since both judge the expected value of $(\beta - R)$ to be greater than 0 for selling prices β of 55 cents or more, Forecaster Two's refusal to sell for a price below 55 cents is a signal to Forecaster One that he is disposed to accept too high a risk of a loss by selling contracts on R for so cheap. Therefore, Forecaster One's selling price should change to agree with Forecaster Two's. This is simply what it means for Forecaster One to change his original selling price expressed by his initial probability of 0.4 for R to the upper probability of 0.55 for R.

The assumption that the currency we are trading is linear is important for pinning down an estimate of an agent's strength of belief in an event occurring, and the operational details of the procedure for eliciting such credences are likewise important for determining whether these numbers are sensible or mere speculative fantasies (Mayo-Wilson and Wheeler 2016). When those conditions are clearly specified and met, and strategic considerations are safe to leave aside, our talk of pricing the value of gambles translates directly to an agent's cognitive epistemic commitments.

Decision theorists did not invent the practice of holding fixed an estimation of another's value of goods to discern what she believes, or vice versa; that trick is as old as humankind. The innovation of mathematical decision theory was to exploit the intimate relationship between belief and value to quantify both the comparison of beliefs and the degree of one's preferences, along a single dimension of value, and to spell out operational procedures for measuring these quantities through manipulations of one to fix a numerical estimate of the other.[19] Philosophers who discuss 'credences' without confronting either their connection to personal preferences or how they are elicited do so at their own peril.

What is novel about the theory of lower previsions, which our lower probability model belongs to, is that it allows an agent to commit to different buying and selling prices for a gamble. The theory of linear previsions, which standard precise Bayesian probability models belong to, does not allow an agent to commit to different buying and selling prices for a gamble but instead takes for granted there is a single number, the agent's fair price. There is nothing imprecise or indeterminate about the highest price you are willing to pay for a gamble or the lowest selling price you are willing to accept for it, regardless of whether those values are different or the same.[20]

In our approach, the span between lower and upper probabilities for a proposition is determined by the range of judgments expressed by a group of peers. As we have argued, there is no reason available to our peers to pick a maximum buying price or a minimum selling price outside this range. But we have also an argument against digging in one's heels. Recall that our discussion of the PIE Principle knocked out conciliatory Bayes responses but left open the option of remaining steadfast in one's opinion, peers be damned. Kelly, for instance, maintains that there isn't enough 'higher-order' evidence from two-person disagreements to warrant either peer to change her view (Kelly 2010). So, a peer who found herself in the situations we are considering should remain steadfast. However, this response confuses the absence of higher-order evidence with the absence of any evidence at all. Put another way, the reason that remaining steadfast is unreasonable is that doing so classifies any information that one acquires through a disagreement as epistemically irrelevant.

To remain steadfast in a peer disagreement is to ignore evidence that one should change her view. Suppose Forecaster One adopts a lower probability of 0.4 and an upper probability of 0.55 for reasons we spelled out above, but peer Two sticks to her guns and persists in viewing 0.55 as her fair price for R. Then, Forecaster Two would discover that Forecaster One refuses to pay more than 40 cents for a unit gamble on R but also refuses to sell gambles to Two for less than 55 cents. What Two learns from One is that One judges the expected value of $(R - \alpha)$ to be *negative* for prices α greater than 40 cents, whereas Two judges her expected loss to remain zero. Conversely, both Two and One judge One's commitments to be non-negative in expectation. So the fallout from this disagreement is that Two receives evidence that she may be exposed to a loss whereas One receives no such evidence. This difference in judgment between One and Two may be defensible if Forecaster Two thought Forecaster One a fool or lacking information that Two had about Rain tomorrow in Riga, but these differences are explicitly ruled out by peer disagreements.[21] The upshot is that by remaining steadfast, Forecaster Two accepts an exposed risk to loss that Forecaster One does not without having a countervailing reason to persist in doing so.

Lastly, our proposal for resolving peer disagreements prescribes a unique set-based credal judgment that all parties to a peer disagreement ought to adopt. Thus, our proposal may be viewed as embracing a central tenet of the *uniqueness thesis* (Feldman 2010; White 2005) while reconciling a seemingly intractable conflict over the nature of the evidence that a peer disagreement generates. For those who

embrace the tripartite distinction between judged true, judged false, and hung-out-in-suspense—supposition still doesn't rate among traditionalists—the unique, conciliatory response to a peer disagreement is to suspend judgment. But this response, given the limited options, saddles you with treating evidence from a peer disagreement as maximally uncertain. For conciliatory Bayesians who restrict themselves to a single determinate probability function, evidence from a peer disagreement is purely ameliorative in character. Our account embraces the insight from traditional suspension-of-judgment views that peer disagreements do not generate ameliorative evidence per se—at least not without some higher-order evidence to tip the scales in favor of some coalition of peers over others.[22] But unlike the naïve suspension of judgment approach, our proposal preserves ranges of agreement and comparative judgments that are lost by naively adopting a partial belief of $\frac{1}{2}$ to represent maximal uncertainty. Finally, unlike both naïve thresholding accounts—which satisfy the Reasonable Range Principle but little more (Foley 1992; Kyburg 2003)—and convex Bayesian accounts, our view emphasizes the basis for group opinion as a repository for information that is common to the group and that may impact how can peers ought to decide to resolve their differences.

Acknowledgements

Notes

[1] For the moment we use the terms 'belief,' 'judgment,' 'view,' and 'opinion' interchangeably to refer to an agent's doxastic attitude toward a particular proposition.

[2] Equally-weighted averaging seems to be a natural belief revision method of the equal-weight view, but the approach has been disputed in the recent literature (Christensen 2011; Kelly 2013). Weighted averaging has a long history in opinion aggregation tracing back to de Finetti (1954) and Stone (1961), who each, independently, proposed weighted averaging to resolve group disagreements among a set of Bayes agents that share the same utility function. Stone used the phrase *opinion pool* to describe this general scenario, and *democratic opinion pool* for the special case when all opinions are equally weighted.

[3] Thanks here to Richard Dawid for pressing us on this point.

[4] A weighted average is non-extreme just in case every peer's opinion takes values in the open interval (0, 1), excluding 0 and 1.

[5] Permissivism is the view that a fixed body of evidence does not necessarily determine a uniquely rational judgment, and there are several varieties of this view in the recent literature (Rosen 2001; Kelly 2010; Douven 2009; Schoenfield 2014; Kopec 2015). In the probabilistic setting, where doxastic judgments are represented by a probability function, a trivial version of permissivism has been acknowledged since Savage's remark that theories of subjective probability "postulate that the individual concerned is in some ways 'reasonable,' but they do not deny the possibility that two reasonable individuals faced with the same evidence may have different degrees of confidence in the truth of the same proposition" (Savage 1954, p. 3). Non-trivial versions of permissivism arise when peers are presumed to share the

same values and same goals of inquiry, where it is a standard assumption in the judgment aggregation and belief pooling literatures to fix such conditions by, for instance, stipulating a single, shared utility function. Because the plausibility of permissivism varies wildly depending both on how one models peer disagreement and how one formulates 'permissivism' in a particular model, a general discussion of permissivism is meaningless.

[6] For example, see White (2005); Elga (2007); Christensen (2007); Christensen (2009); Kelly (2010); Feldman (2010); Ballantyne and Coffman (2011); Schoenfield (2014); Levinstein (2015); Russell, Hawthorne, and Buchak (2015).

[7] For discussions on high-order evidence in peer disagreements, see Christensen (2010); Kelly (2010); Lasonen-Aarnio (2014).

[8] For example, to verify the first row of Table 1, $p_1(R \wedge H) = p_1(R)p_1(H) = (0.4)(0.2) = 0.08$, and $p_2(R \wedge H) = p_2(R)p_2(H) = (0.55)(0.6) = 0.33.$, yet $p^*(R \wedge H) = \frac{p_1(R)p_1(H) + p_2(R)p_2(H)}{2} = 0.205 \neq 0.19 = \frac{p_1(R) + p_2(R)}{2} \times \frac{p_1(H) + p_2(H)}{2} = p^*(R)p^*(H)$.

[9] Our argument is a variation of one that Henry Kyburg and Michael Pittarelli (1996) make against Levi's E-admissibility decision rule, which, in Levi's original form, presupposes non-extreme weighted averaging. Also, for the sake of the argument, we assume throughout that the utility of money for peers is linear.

[10] That is, €18.75(0.205) − €81.25 (0.195) + €143.75 (0.27) − €81.25 (0.33) = 0

[11] That is, €18.75(0.19) − €81.25 (0.21) + €143.75 (0.285) − €81.25 (0.315) ≈ €1.88.

[12] See Stewart and Ojea Quintana (2015) for an excellent review of Bayesian pooling methods and their properties, and Wheeler (2012) for an objection to Williamson's Objective Bayesian approach.

[13] Thanks to Jennifer Carr for raising this objection to us.

[14] Although irrelevance, epistemic independence, and stochastic independence (factorization) are logically equivalent for a single probability measure, modulo some regularization condition to avoid conditioning on zero probability events, these three concepts are logically distinct for lower and upper probabilities. See Pedersen and Wheeler (2014) for examples and discussion.

[15] There is a fascinating literature exploring structural judgments under $\underline{P}$, including the plurality independence concepts (Couso, Moral, and Walley 1999; de Cooman, Miranda, and Zaffalon 2011; Cozman 2012; Pedersen and Wheeler 2014; Augustin, Coolen, de Cooman, and Troffaes 2014) and structural judgments, and the differences between permutability and exchangeability (Walley 1991; de Cooman and Miranda 2007).

[16] That is, replace our (finite) $\mathbb{P}$ by the set of probability measures constructed by all possible linear weighted averages of p_1 and p_2, that is $\mathrm{co}\,\mathbb{P} = \{p' : p' = \lambda p_1 + (1 - \lambda)p_2, \text{ for all } 0 \leq \lambda \leq 1\}$.

[17] Compare with Joyce (2010, p. 287).

[18] This assumption can be relaxed, allowing us to start with some or all agents having credal commitments that are indeterminate or to consider iterative peer disagreements that start with a group of standard Bayes agents but where indeterminacy is introduced by the resolution of a sequence of disagreements. We may even dispense with probabilities altogether and give a general qualitative account in terms of desirable gambles (Williams 1975; Walley 2000). Each is beyond the scope of this paper.

[19] Viewing the value of goods along a single ratio scale does not come automatically. See Elizabeth Anderson's (1987) persuasive arguments for the heterogeneity of values. Consequences for epistemic decision theory are discussed in Mayo-Wilson and Wheeler (2016).

[20] The term 'imprecise credences' was coined relatively recently as a broad shorthand for some or another *no-fair-price* attitude that calls on, or calls out, imprecise probability theory. This slogan, which has spread like kudzu, is now a source of considerable confusion. If there is a clear interpretation of probability running in the background, then imprecise credal talk that slides between psychological states, observable behavior, mathematical properties, or what have you, is a manageable affair: the leeway afforded by natural language is sometimes an ally in getting our ideas across. But such carelessness must be earned. For without the backbone provided by a clear interpretation, the term 'imprecise credences' is a recipe for mushy thinking about imprecise probabilities.

[21] And our assumptions about a shared linear scale of value rule out cases of different attitudes toward risk.

22 Scott Sturgeon (2010) and Haenni et al. (2011) each consider interpreting the span between a lower and upper probability the degree to which an agent suspends judgment.

References

Anderson, E. S. (1987). *Value in Ethics and Economics*. Ph. D. thesis, Harvard University, Cambridge, MA.

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

Ballantyne, N. and E. Coffman (2011). Uniqueness, evidence, and rationality. *Philosophers' Imprint 11*(18), 1–13.

Bernoulli, J. (1713). *Ars Conjectandi*. Basel: Thurnisius.

Boole, G. (1854). *An Investigation of the Laws of Thought*. New York: Dover.

Bradley, S. (2014). Imprecise probabilities. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 ed.). CSLI Publications.

Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review 116*, 187–217.

Christensen, D. (2009). The epistemology of controversy. *Philosophy Compass 4*(5), 756–767.

Christensen, D. (2010). Higher-order evidence. *Philosophy and Phenomenological Research 81*, 185–215.

Christensen, D. (2011). Disagreement, question-begging and epistemic self-criticism. *Philosophers' Imprint 11*(6), 1–22.

Couso, I., S. Moral, and P. Walley (1999). Examples of independence for imprecise probabilities. In G. de Cooman (Ed.), *Proceedings of the First Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Ghent: ISIPTA.

Cozman, F. (2012). Sets of probability distributions, independence, and convexity. *Synthese 186*(2), 577–600.

de Cooman, G. and E. Miranda (2007). Symmetry of models versus models of symmetry. In W. Harper and G. Wheeler (Eds.), *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* London: King's College Publications.

de Cooman, G., E. Miranda, and M. Zaffalon (2011). Independent natural extension. *Artificial Intelligence 175*, 1911–1950.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré 7*(1), 1–68. Trans. Henry E. Kyburg, Jr., Foresight: Its logical laws, its subjective sources. In H. E. Kyburg, Jr. and H. Smokler (Eds.), *Studies in Subjective Probability*. New York: John Wiley and Sons, 1964.

de Finetti, B. (1954). Medi di decisioni e media di opinioni. *Bolletino dell'Istituto internazionale di Statistica 34*, 144–157.

de Finetti, B. (1974). *Theory of Probability* (1990 ed.), Volumes I and II. New York: John Wiley.

Douven, I. (2009). Uniqueness revisited. *American Philosophical Quarterly 46*(4), 347–61.

Douven, I. (2010). Simulating peer disagreements. *Studies in History and Philosophy of Science 41*, 148–157.

Elga, A. (2007). Reflection and disagreement. *Noûs 41*, 478–502.

Feldman, R. (2010). Reasonable religious disagreements. In A. Goldman and D. Whitcomb (Eds.), *Social Epistemology: Essential Readings*. New York: Oxford University Press.

Foley, R. (1992). The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly 29*, 111–21.

Haenni, R. J. W. Romeijn, G. Wheeler, and J. Williamson (2011). *Probabilistic Logics and Probabilistic Networks*. Synthese Library. Dordrecht: Springer.

Halmos, P. R. (1950). *Measure Theory*. New York: Van Nostrand Reinhold Company.

Joyce, J. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives 24*(1), 281–323.

Kelly, T. (2010). Peer disagreement and higher order evidence. In A. Goldman and D. Whitcomb (Eds.), *Social epistemology: Essential readings*. New York: Oxford University Press.

Kelly, T. (2013). Disagreement and the burdens of judgment. In D. Christensen and J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

Koopman, B. O. (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics 41*(2), 269–292.

Kopec, M. (2015). A counterexample to the uniqueness thesis. *Philosophia 43*(2), 403–409.

Kyburg, Jr., H. E. (2003). Are there degrees of belief? *Journal of Applied Logic 1*, 139–149.

Kyburg, Jr., H. E. and M. Pittarelli (1996). Set-based Bayesianism. *IEEE Transactions on Systems, Man and Cybernetics A 26*(3), 324–339.

Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research 88*, 314–345.

Lehrer, K. and C. Wagner (1983). Probability amalgamation and the independence issue: A reply to Laddaga. *Synthese 55*, 339–346.

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

Levi, I. (1990). *Hard Choices: Decision making under unresolved conflict*. Cambridge: Cambridge University Press.

Levinstein, B. (2015). Permissive rationality and sensitivity. *Philosophy and Phenomenological Research*.

Mayo-Wilson, C. and G. Wheeler (2016). Epistemic decision theory's reckoning. Unpublished Manuscript.

Miranda, E. and G. de Cooman (2014). Structural judgements. In T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes (Eds.), *Introduction to Imprecise Probabilities*, Probability and Statistics. West Sussex: Wiley and Sons.

Moss, S. (2015). Time-slice epistemology and action under indeterminacy. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Volume 5. Oxford: Oxford University Press.

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

Ramsey, F. P. (1926). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays*, 1931. London: Kegan, Paul, Trench & Company.

Rosen, G. (2001). Nominalism, naturalism, epistemic relativism. *Philosophical Perspectives 15*, 69–91.

Russell, J. S., J. Hawthorne, and L. Buchak (2015). Groupthink. *Philosophical Studies 172*(5), 1287–1309.

Savage, L. J. (1954). *Foundations of Statistics*. New York: Wiley.

Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs 48*(2), 193–218.

Smith, C. A. B. (1961). Consistency in statistical inference (with discussion). *Journal of the Royal Statistical Society 23*, 1–37.

Stewart, R. and I. Ojea Quintana (2015). Probabilistic opinion pooling with imprecise probabilities. Unpublished manuscript.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics 32*(4), 1339–1342.

Sturgeon, S. (2010). Confidence and coarse-grain attitudes. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Volume 3. Oxford University Press.

Walley, P. (1981). Coherent lower (and upper) probabilities. Statistics Research Report 22, University of Warwick, Coventry, England.

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

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

Wheeler, G. (2012). Objective Bayesianism and the problem of non-convex evidence. *The British Journal for the Philosophy of Science 63*(3), 841–50.

White, R. (2005). Epistemic permissiveness. *Philosophical Perspectives 19*(1), 445–459.

Williams, P. M. (1975). Notes on conditional previsions. School of Mathematical and Physical Sciences, University of Sussex. Republished in *International Journal of Approximate Reasoning*, *44*(3): 366–383, 2007.

Williamson, J. (2015). Deliberation, judgment and the nature of evidence. *Economics and Philosophy 31*(1), 27–65.