Confidence in Belief, Weight of Evidence and Uncertainty Reporting

Brian Hill　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　HILL@HEC.FR
GREGHEC, CNRS & HEC Paris, Jouy-en-Josas, France

Abstract

One way of putting a popular objection to the Bayesian account of belief and decision is that it does not allow a role for something akin to the Keynesian concept of 'weight of evidence' in choice. This paper argues that a recently-defended approach, which refines the credal-set representation of beliefs to give pride of place to an agent's confidence in her beliefs, can do so fruitfully. Motivated by the use of confidence by the IPCC and US Defense Intelligence Agency in their assessments of uncertainty, the paper then considers the consequences of the proposed approach for uncertainty reporting. On the one hand, when connected to decision, the model affords a clear separation of the belief and value factors: an important quality in policy making contexts where these are the responsibility of different actors. On the other hand, the issue of inter-agent confidence calibration is discussed, and a calibration scale is proposed and defended, on the basis of weight-of-evidence version of David Lewis's Principal Principle.

Keywords: confidence in belief, uncertainty reporting, Keynesian weight of evidence, rational decision, confidence ranking, credal sets.

1. Introduction

The Bayesian representation of beliefs by probability measures is the centre-piece of an apparatus central in decision theory, statistics, decision analysis and uncertainty communication. Sets of probability measures—sometimes called *credal sets* or *imprecise probabilities*—have been defended as an alternative representation of rational belief, which overcomes some of the most troublesome challenges to Bayesianism [34, 35].

A major motivation focusses on Bayesianism's purported inability to faithfully capture apparently relevant properties of beliefs. Consider two urns each containing 100 balls, each of which is only black or white: for one of the urns (the small-sample urn), you have observed 2 draws (with replacement), one of which was black; for the other (the large-sample urn), you have observed 1 million draws (with replacement), half of which were black. Bayesianism enjoins you to have a precise degree of belief (or credence, or subjective probability) about the colour of the next ball drawn, for each urn—say, $\frac{1}{2}$ in it being black for both urns. However, as Keynes noted [32], even if the *balance of evid-*

ence in favour or against the next draw being black is the same in both cases, the amount of or *weight of evidence* differs. So even if you were willing to adopt the degree of belief $\frac{1}{2}$ for both events, it seems clear that the weight of evidence in support of that (degree of) belief is larger in the case of the large-sample urn. If the balance of evidence is something to be reflected in a rational agent's degree of belief, it is difficult to see why the weight should not.

Indeed, some Bayesians suggest that weight of evidence can be captured using the apparatus of (precise) probability measures alone, in a variety of ways [19, 45, 30]. However, when it comes to the decision of which urn to bet on, Bayesianism dictates that, since the degrees of belief are the same, you should be indifferent between betting on the colour of the next ball drawn from the small- or large-sample urns. Whatever ways there are of capturing something like weight in the framework of Bayesian probability, they remain irrelevant for choice. As noted by Ellsberg [11], this seems both descriptively inaccurate and normatively unreasonable. Many people strictly prefer betting on the colour from the large-sampled urn [1], and indeed, the difference in your ignorance about the two urns—or in the weight of evidence in the two cases—seems to justify a preference between them [35, 18]. In a word, the objection is that something of the order of the weight of evidence can reasonably have a role in choice, and Bayesianism denies it any such role.

This stylized example is indicative of real-world challenges. A climate scientist considering his judgement on a climate prediction in a particular emissions scenario will have to take into account research in several fields with a variety of methods. In doing so, she will have to make calls about the direction in which that evidence is pointing, but also about the quality and amount of the existing evidence. These will be 'subjective' elements of her belief, like Bayesian subjective probabilities, and such domains make use of belief elicitation from experts, or 'expert elicitation' [9, 40, 48]. However, unlike Bayesian beliefs, there seem to be two factors—one the subjective equivalent of something like the balance, and the other the subjective equivalent of the weight. It seems reasonable that the latter may have a role to play in rational decision, despite the fact that, as the previous example shows, Bayesianism denies it any such role. Henceforth, we shall refer to the probability judgements—reflecting the direction evidence is pointing—as the *beliefs* or *credal judgements*, and use the

term *confidence in beliefs* to denote the subjective appraisal of the support for them.

Such factors are already appearing in *uncertainty reporting*. Institutions such as the IPCC use an uncertainty reporting language that permits expression both of probability judgements concerning events of interest, as well as some indication of the quality and amount of evidence *inter alia* behind those judgements, usually called *confidence* [38]. For instance, here is a statement from a recent report:

> In the Northern Hemisphere, 1983–2012 was *likely* [i.e. probability between 66% and 100%] the warmest 30-year period of the last 1400 years (*medium confidence*). [29, 3][1]

A similar language is used by the US Defense Intelligence Agency [13], which also faces situations of severe uncertainty.

This paper considers how something of the order of weight of evidence or confidence in beliefs can be formally represented, and how such models can contribute to effective uncertainty reporting. It first presents an account of confidence in beliefs and its role in decision recently developed in the economics and philosophy literature. It then examines two theoretical desiderata for an uncertainty reporting language, showing how the proposed approach satisfies one, and proposing a calibration scale allowing it to satisfy the other.

2. Confidence in Beliefs and Decision

2.1. A Model of Confidence in Beliefs

Let us begin by recalling the imprecise probabilities (IP) or 'set of probability measures' representation of belief. Let Δ be the set of probability measures over a fixed state space Ω with σ-algebra Σ. Statements about degrees of belief—such as '*A* has a higher degree of belief than *B*', '*A* has a higher degree of belief than $\frac{1}{2}$', '*A* is probabilistically independent of *B*' and so on— are called *credal statements*, or *credal judgements*, and correspond to sets $\mathscr{P} \subseteq \Delta$. ('*A* has a higher degree of belief than *B*' is the set $\{p \in \Delta : \ p(A) \geq p(B)\}$ and so on.) According to the IP representation (defended by [35, 31], for example), an individual's state of belief is represented not by a single probability measure but by a set $\mathscr{C} \subseteq \Delta$ of such measures, which we call an *imprecise probability (IP)* or *credal set*. For example, an agent will have a higher degree of belief for a proposition *A* than *B* if $p(A) \geq p(B)$ for all probability measures $p \in \mathscr{C}$. More generally, an agent will *hold a credal judgement P* if $\mathscr{C} \subseteq P$. The set of probability measures $\mathscr{C}$ involved in the imprecise

probability representation hence generates a set of credal judgements, namely those to which the agent adheres.

If this representation is to properly overcome the challenge of appropriately capturing confidence or weight of evidence, credal sets need to be construed as reflecting such factors. Some have suggested considering them to reflect both belief (or balance of evidence) and confidence (or weight of evidence) [50]. For instance, in the leading example, an agent may hold the (precise) credal judgement of 0.5 for black being drawn from the large-sample urn, but not hold this judgement for the small-sample urn. There will be other, more imprecise, credal judgements (e.g. $[0.01, 0.99]$) held by the agent concerning both the small- and large-sample urns. Precisely which judgements are held for, say, the small-sample urn will depend on how the agent trades off considerations of balance (which points to 0.5) and weight (which pushes for larger intervals).

Note that this approach has trouble accounting for the intuition that one can have the same precise credal judgement of 0.5 for both urns, just with much less confidence—or lower weight of evidence—in one case. This is because the IP approach is an all-or-nothing affair: credal judgements are either held or not, so the only way to distinguish the two cases is to jettison judgements that a typical agent would seem to hold, but with low confidence. This may be an indication that the approach has trouble reflecting the fact that confidence—like weight of evidence—comes in degrees.

A recently-defended belief representation, which captures specifically such degrees, replaces the single set of probability measures by a *nested family* of such sets: that is, a family where each member is contained in or contains each other member [24, 27].[2] Such a nested family is called a *confidence ranking*. The sets in the family correspond to levels of confidence, with larger sets corresponding to higher levels (see Figure 1). As noted above, each set generates a collection of credal statements: these represent the credal judgements the agent holds to the corresponding level of confidence. For larger sets in the family, corresponding to higher confidence levels, the generated collections of credal statements are smaller, and so fewer credal judgements are held by the agent at higher confidence levels. In particular, the nested structure implies that any credal judgement held with high confidence is retained at any lower confidence level, as one would expect. See [24, 27] for more detailed discussion.

For the purposes of this paper, the sets in the confidence rankings are not assumed to be closed or convex, and the rankings themselves need not be continuous. (In decision-theoretic representation results, such properties often drop out naturally, see [24, 26].) The interpretative remarks above and the points made below do not dependent

---

1. Note that the event in question is non-stochastic (it is either true or false), so the probability judgement cannot be interpreted in an objective or frequentist way, and is most naturally construed as epistemic or subjective probability.

2. Similar representations have been proposed and explored by [14, 41]. A crucial difference is in the ordinality of the model used here at the second order level; see [27] for further discussion.
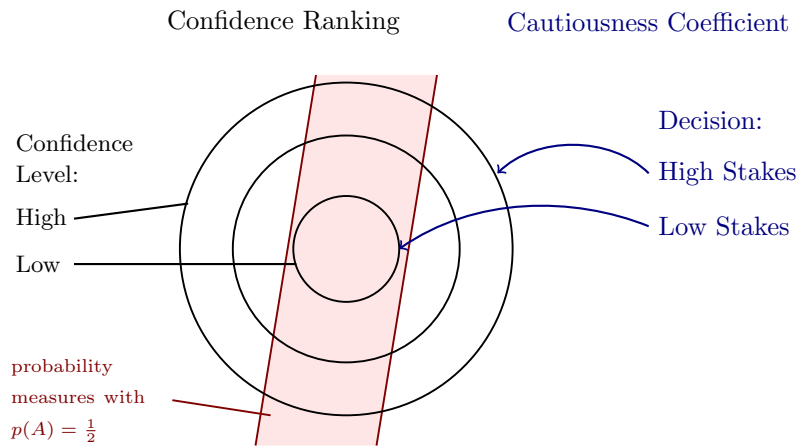
Figure 1: Representation of confidence in beliefs (black) and relation to decision (blue)

on such properties. We do assume that confidence levels are totally ordered, as in the IPCC scale. The framework can be extended beyond this assumption; see [23] for a preliminary discussion.

Just as credal sets correspond to sets of credal judgements, each confidence ranking Ξ induces an order $\succeq_\Xi$ on credal statements—defined by $P \succeq_\Xi Q$ if and only if, for all $\mathscr{C} \in \Xi$, $\mathscr{C} \subseteq Q$ implies $\mathscr{C} \subseteq P$—which captures the relative confidence that the agent has in them. So, for example, if $p(A) = \frac{1}{2}$ for all probability measures in a small set in the confidence ranking Ξ, but not in larger sets, but $p(B) = \frac{1}{2}$ for all probability functions in a larger set in Ξ, then Ξ represents an agent who is more confident in her assessment of $\frac{1}{2}$ for her degree of belief in B than in her assessment of $\frac{1}{2}$ for her degree of belief in A (see Figure 1 for a graphical representation of the confidence in judgements concerning A). Taking as A the event that the next ball drawn from the small-sample urn is black, and similarly for B and the large-sample urn, Ξ thus faithfully renders the intuition that the credence of $\frac{1}{2}$ is held concerning both urns, but with different amounts of confidence. It can thus reflect the differing weights of the evidence supporting these judgements.

Note furthermore that whilst such an agent will have low confidence for the credence $\frac{1}{2}$ in the case of the small-sample urn, she will have higher confidence for less precise credal judgements. For instance, she will be maximally confident in the credal judgement $[0.01, 0.99]$ concerning that urn if $p(A) \in [0.01, 0.99]$ for all probability functions contained in some set in Ξ. So confidence rankings explicitly portray the trade-off between credal precision and confidence (or balance precision and weight): more precise credal judgements may be held, but with less confidence (or support from the evidence). Whilst set out explicitly, the agent is not assumed or asked to make a call about what

confidence gain justifies a particular loss in precision: she is need not settle on a single credal set.[3]

2.2. Confidence in Belief and Decision Making

The previous model has been developed and defended in tandem with an account of the role of confidence in belief in decision making, based on the following maxim:

Maxim the higher the stakes involved in the decision, the more confidence is required in a belief for it to play a role.

This appears to be a sensible way of relating two aspects of a decision: its importance (or the stakes involved in it), and the beliefs one relies on to take it. To the extent that confidence reflects something of the order of the weight of evidence behind a belief or credal judgement, the maxim sets out a role for weight: *ceteris paribus*, more important decisions should be based on beliefs benefiting from more support or weight of evidence. It directly motivates the following formal framework for decision.

Assume firstly that to each decision or option the agent is faced with, she can associate a level of confidence appropriate for it. As noted above, the confidence levels correspond to sets in the confidence ranking: so assigning a confidence level to a decision amounts to assigning a set in the confidence ranking. Moreover, the maxim requires that the assignment is made on the basis of the stakes involved: more important decisions—those involving larger stakes—call for more confidence, and are thus associated to higher confidence levels, which correspond to larger sets in the confidence ranking. In summary then, we take

---

3. There are interesting synergies between this account of weight of evidence and Isaac Levi's reading of Keynes in [36]. For instance, he assimilates Keynes's weight of argument with Shackle's potential surprise [44], which corresponds roughly to an order on the state space. Somewhat analogously, the confidence rankings proposed here are orders on the probability space (and hence arguably capture more appropriately the balance precision-weight trade-off).

a function D that assigns a set in the confidence ranking to each decision, such that decisions with higher stakes are sent to larger sets (see Figure 1). Such a function is called a *cautiousness coefficient*. It represents the agent's attitudes to uncertainty, in much the same way as the utility function in standard Bayesianism is often interpreted as a representation of her desires for outcomes (Section 3.1).

The suggested rendition of the aforementioned maxim is simple: to evaluate an option, use the set of probability measures in the confidence ranking that corresponds to the decision at hand according to the cautiousness coefficient. Why? This amounts to using the credal judgements held to the corresponding level of confidence. But this is the level picked out (by the cautiousness coefficient) as being appropriate for the decision at hand, on the basis of the stakes involved. So using this set of probability measures basically means that the agent only relies on beliefs that she holds with enough confidence given the stakes involved in the decision. This procedure is thus faithful to the maxim.

This does not yield a single decision rule as much as a family of rules. Indeed, it just picks out a set of probability measures, or IP, but does not specify how to choose on the basis of it. Several decision rules for IPs have been proposed; each of these, when inserted into the framework, will result in a corresponding confidence-based decision rule. For example, using the maximin-EU or Γ-Maximin decision rule [17, 3], which looks at the lowest expected utility calculated across the set of probability measures, naturally yields a rule which evaluates an act f according to:

$$\inf_{p \in D(f)} EU_p f \qquad (1)$$

where $EU_p f$ is the expected utility of f calculated with probability p and utility U,[4] and D is a cautiousness coefficient assigning to every act a confidence level. Alternatively, if one uses the unanimity or maximality rule [4, 49], then f will be chosen over g if and only if:[5]

$$EU_p f > EU_p g \qquad \text{for all } p \in D((f,g)) \qquad (2)$$

where D is a cautiousness coefficient assigning to every binary choice (pairs of acts) a confidence level. [6]

These examples also illustrate how confidence-based decision rules are basically extensions of (corresponding) IP decision rules. For example, the standard maximin-EU rule is just like (1) except that $D(f)$ in the infimum is

replaced by a fixed set $\mathscr{C}$, and similarly for (2).[7] So the confidence-based family of rules can account for any choice patterns that imprecise probabilities can. Returning to the previous example, at any reasonable confidence level, the agent will endorse the credal statement that the probability of getting black from the large-sample urn is $\frac{1}{2}$; by contrast, whenever the confidence level is high enough, she may not hold such a precise credal judgement concerning the small-sample urn, instead restricting herself to intervals, such as $[\frac{1}{4}, \frac{3}{4}]$. Under (1), when the stakes are high enough to merit such a confidence level, she will use $\frac{1}{2}$ to evaluate the act of betting on the large-sample urn, whilst look at the minimal expected utility over $[\frac{1}{4}, \frac{3}{4}]$ in evaluating the bet on the small-sample urn. Since the latter value is lower than the former, she will prefer to bet on the large-sample urn. To the extent that confidence can reflect the weight of evidence behind a credal judgement, this captures faithfully the intuitive role of weight of evidence in choice noted in the Introduction.

Analysis of several decision models in the confidence family points to a single essential behavioural difference between the imprecise probability and the confidence-based approaches: *stakes independence*. Formulated in the language of bets, stakes independence is the assumption that, if an agent is willing to buy a bet on event E with stakes S[8] for a price of $q.S$, then she is willing to buy a bet on the same event with stakes S' for a price of $q.S'$, and this for any S and S'. In other words, the lower betting quotient—the supremum proportion of the stakes for which the agent is willing to buy the bet—is independent of the stakes.[9]

The rational credentials of this assumption are questionable. Consider an urn containing 100 black and white balls, and an agent who knows that there are at least 10 black balls in it and has observed two draws (with replacement), one of which is black. She may be willing to buy a \$10 million bet on the next ball drawn being black for \$1 million, but may be willing to pay much more than \$0.10 to buy a bet on the same event when only \$1 is at stake. After all, in latter case, less is at stake, so it does not seem so unreasonable to rely on beliefs in which she has more limited confidence in evaluating the gambles. This would suggest that the lower betting quotients may depend on the stakes involved. Certainly, such dependence does not appear to be irrational. Moreover, it naturally seems to go in a particular direction: when the stakes are higher, the decision maker may reasonably refuse to buy at betting quotients that she would have accepted at lower stakes.

---

4. Since the focus here is on belief, we follow standard Bayesianism in assuming throughout the paper a precise utility or desirability function as a representation of desires over outcomes.

5. There are different versions of this rule depending on the sort of dominance required, e.g. strict or weak order in (2). The points made in the discussion hold for all variants.

6. As the two examples illustrate, models in the confidence family may also differ on their treatment of the stakes associated to a decision. See [24, 26, 25] for technical details and further discussion.

---

7. The maximin-EU rule [17] is standardly formulated using convex closed sets and hence minima instead of infima; the corresponding confidence representation involves closed and convex confidence rankings [24].

8. I.e. the bet pays S is E and nothing if not.

9. Technically, in terms of lower previsions, stakes independence is equivalent to the positive homogeneity axiom [49].

It turns out, in a range of frameworks and under a variety of decision models, that replacing a stakes-independence condition in an appropriate characterisation of imprecise probabilities by a stakes-dependence condition along these lines yields a characterisation of the confidence model [24, 26, 27]. To the extent that stakes independence appears to be overly restrictive as a normative condition, whilst stakes dependence does not, this might suggest that the confidence approach has as strong, if not stronger, normative credentials when it comes to decision than the standard imprecise probability approach.

Moreover, this suggests a 'rule-of-thumb' way of understanding confidence in beliefs in its relationship to choice. The interpretation of degrees of belief in terms of betting quotients gives a useful grasp on the concept, which can help guide intuition. The previous discussion suggests a similar 'proxy' for confidence: the confidence in a degree of belief is reflected in the stakes to which one is willing to let that degree of belief guide one's betting behaviour. As such, the introduction of confidence in belief appears a natural addition to degrees of belief: beyond the odds one gets (reflecting degrees of belief), there is the issue of how much one is willing to bet on those odds (reflecting confidence in those beliefs).

Note finally that the combination of this account of decision with the connection between weight of evidence and confidence provides a reply to the question of the role that weight of evidence should play in choice, which Keynes himself famously found 'highly perplexing' [32, p357].[10] Weight of evidence—and more generally confidence—regulates the probability judgements which can be relied upon, according to the maxim that for more important decisions, only those in which one has more confidence are appropriate for use.

3. Confidence and Uncertainty Reporting

The ubiquity of the Bayesian model of belief in practical applications is at least in part due to its capacity to support group decision making, especially in situations where different actors (experts, scientists, the public, policy makers) are responsable for different factors or parts of the decision making process ('group beliefs', 'group values'). Two properties of the model are theoretically central in such situations. One is the clear separation of the elements of a decision paradigm which correspond to or capture beliefs—the probabilities, in the Bayesian approach—and those which reflect values or tastes—the utility function. This permits the theoretical division of decision-making

labour: experts or scientists are responsible for the probabilities; the utilities or values should be those of the policy makers or society. The other is a clear language for uncertainty communication: the language of probabilities is one that, in principle, is fully understood by all parties. When an expert reports the probability of a major natural disaster in the next 20 years in New York as 20%, policy makers (in theory) understand that assessment the same way as the expert.

Of course, the picture is not so rosy in practice: the fact-value distinction is often portrayed as more fragile as it would be in principle under the Bayesian framework [10, 47], and communication of probabilities is rarely as faithful and unambiguous as theory would suggest [7, 8, 6]. However, any account of uncertainty capable of supporting real decision making should at least possess these characteristics *in theory*. We now consider how the confidence approach fairs on each.

3.1. Belief-Taste Separation

Compared to Bayesianism, the confidence-based approach involves two novel elements: the confidence ranking and the cautiousness coefficient.[11] As explained in Section 2.1, the confidence ranking captures the decision maker's state of belief, incorporating in particular her confidence in her beliefs. As for the cautiousness coefficient, it can be understood as a representation of her attitude to choosing in face of limited confidence. This interpretation is suggested by its role in the model. It involves a judgement as to the appropriate confidence level for the decision at hand, and hence reflects the extent to which the decision maker is willing to rely on beliefs held with limited confidence in such a decision. Ann and Bob are each evaluating the bet, with stakes of \$1 billion, on black from an urn about which no information has been provided except that it only contains black and white balls, and suppose that they have the same confidence ranking. Suppose that Ann's cautiousness coefficient assigns this decision to the set $\mathscr{C}_0 = \{p : 0 \leq p(black) \leq 1\}$ in their confidence ranking, whereas Bob's assigns it to the smaller set $\mathscr{C}_1 = \{p : 0.25 \leq p(black) \leq 0.75\}$. Since $\mathscr{C}_1$ is in Ann's confidence ranking, it represents beliefs that she holds; however, she feels uncomfortable relying on beliefs held with that level of confidence in such a high-stakes decision. Bob, by contrast, is less averse to mobilising beliefs held with this much confidence in decisions of such importance. If you will, he is readier to take the 'epistemic risk' of relying on beliefs held with limited confidence when the stakes are so high. Ann and Bob differ in their attitudes, or tastes, for choosing on the basis of beliefs held with limited confidence.

---

10. 'In marginal cases, therefore, the coefficients of weight and risk as well as that of probability are relevant to our conclusion [concerning the preferable course of action]. It seems natural to suppose that they should exert some influence in other cases also, the only difficulty in this being the lack of any principle for the calculation of the degree of their influence.' [32, p360]

11. The final element is the utility function, which, as standard, can be interpreted as reflecting the decision maker's desires for outcomes, and hence deserves no further discussion here.

The important point is that the cautiousness coefficient is conative in character: it reflects a taste or value judgement, rather than something of the order of a belief. The model thus neatly separates the doxastic element—fully captured by the confidence ranking—from conative attitudes—reflected entirely by the utility function and the cautiousness coefficient.

The economic literature on decision suggests that such a clean separation of doxastic and conative attitudes is generally lacking from decision models built on IPs. Consider for illustration the maximin-EU rule, which evaluates an act f according to:

$$\inf_{p \in \mathscr{C}} EU_p f \qquad (3)$$

where $\mathscr{C}$ is a set of probability measures (and the rest of the notation is as in Section 2.2). A tempting, and perhaps even popular interpretation of $\mathscr{C}$ is as representing the decision maker's state of belief: after all, it seems to be the equivalent in this model of the Bayesian subjective probability. Now consider the 'unknown' urn from above: on the basis of the 'objective' information available, any composition of the urn is possible, so the information is summarised by a set of probability measures $\mathscr{C}_0 = \{p : 0 \le p(black) \le 1\}$. Suppose that Ann's preferences can be represented by this set, whereas Bob's preferences adhere to the maximin-EU rule, but with $\mathscr{C}_1 = \{p : 0.25 \le p(black) \le 0.75\}$. Where do these agents differ: in their beliefs or in their tastes? On the one hand, it could be that Bob has further beliefs, beyond the available information, that allow him to restrict the set of probability measures. On the other hand, Bob could have a greater tolerance of uncertainty—or less cautious attitude. Indeed, he will be more ready than Ann to choose an uncertain option over one that involves no uncertainty—behaviour which is standardly taken to betray less uncertainty aversion on his part [15, 16], where uncertainty aversion, like the sister-notion of risk aversion [42, 2], is taken to reflect decision makers' tastes for bearing uncertainty. The use of IPs with the maximin-EU model is not rich enough to decide the question—or, indeed, to represent the difference between these two possibilities. One often concludes that there is no clear interpretation of the set $\mathscr{C}$: it reflects aspects of both belief and attitude to or taste for uncertainty [33].

By contrast, the confidence model can deal with this issue quite straightforwardly. If Ann and Bob have different confidence rankings—if, in particular $\mathscr{C}_1$ does not belong to Ann's confidence ranking—then they have different beliefs. If, on the other hand, they have the same confidence ranking, but different cautiousness coefficients—with Ann's picking out $\mathscr{C}_0$ for a particular decision and Bob's $\mathscr{C}_1$— then their tastes or values for choosing on the basis of limited confidence are different. Just as so-called 'comparative statics' results on the maximin-EU model and other IP models in the economic literature tend to point up the ambiguity in the interpretation of the single set of priors

[15, 16], corresponding results on confidence models corroborate the clear distinction between the belief and taste elements of the representation [24, 26].[12]

These issues for IP models could be seen as indicative of deeper, interrelated problems, concerning uncertainty communication and incorporation of evidence. For instance, since the set of probability measures *can* reflect the decision maker's attitude to uncertainty, how are we sure, when an agent reports a set of probability measures in good faith for use to guide choice in the context of such a rule, that she is not inadvertently letting her tastes for uncertainty contaminate her report, and the subsequent choice? This echoes concerns raised in the literature on (experimental) elicitation of imprecise probabilities [46, 51, 52]. Reporting probability intervals requires subjects to trade-off between the accuracy of the estimate and its informativeness, so the interpretation of any intervals elicited depends on how subjects make these trade-offs. To the extent that they may involve value judgements (as to whether it is better to be more precise but wrong, or not, for the decision in hand), this is basically a consequence of the lack of a clear separation between doxastic and conative attitudes.

In practice, the sorts of trade-offs just mentioned are often related to issues raised in Section 2.1 concerning the incorporation of evidence. More imprecise—and less informative—credal judgements may benefit from more support or weight of evidence—and hence be more accurate. This suggests that the accuracy-informativeness trade-off is related to the balance precision-weight one discussed above. The problem is that such trade-offs inevitably involve considerations of which—accuracy or informativeness, precision or weight—is more important, and that brings in value considerations that, moreover, may depend on the (decision) context. Weight of evidence can only be incorporated into IPs by trade-offs incorporating values as well, hence forsaking any clear belief-taste separation, of the sort important for group decision making.

By contrast, the confidence approach can comfortably separate weight of evidence—or more generally confidence—from the precision of credal judgements. In the aforementioned accuracy-informativeness trade-off, the confidence ranking faithfully represents the solidity of different credal judgements, but without actually relying on any of the trade-offs actually being made. Perhaps relatedly, the approach separates this from value considerations about how much solidity, weight or confidence is appropriate for a given situation, with the cautiousness coefficient regulat-

---

12. We focus here on the separation between beliefs and tastes for choosing on the basis of limited confidence—or 'caution in decision'—at a given moment. Of course, a full picture would require an account of update of confidence in beliefs. Such an account is provided in [28], separating in particular the previous elements from the agent's inclination to jump to strong conclusions on the basis of limited data—or 'caution in learning'.

ing the extent to which the agent is prepared to sacrifice confidence for a sharper set of beliefs to guide decision.

The confidence framework thus offers a clear story about its central elements. As such, it provides simple guidance for its use in public decision making: one should look to the experts to provide the confidence ranking, and to the policy maker to fix the cautiousness coefficient. In so doing, the experts would not be asked to decide on any trade-offs, but only report their credal judgements and the confidence in them—including the weight of evidence supporting them—while it would fall to the policy maker to decide on how much confidence is appropriate for the decision at hand. As noted, some institutional actors currently use a reporting framework involving the possibility of expressing confidence over probability judgements; [5, 22] relate it to the confidence approach, bringing out the implications and recommendations for practice going forward.

3.2. Communication and Calibration

In the previous model, confidence in credal judgements is ordinal (Section 2.1). This promises to facilitate the use of the model in situations where confidence assessments need to be provided by experts [40]. In such cases, only ordinal confidence judgements are required from experts: they need only say *whether* they are more confident in one credal judgement than another, but not by *how much*. On the other hand, it poses a challenge for communication, where, beyond the confidence ordering of credal judgements, one needs to relate confidence levels across actors. How can agents know that they are talking about the same level of confidence? This problem, which has been noted for various non-Bayesian uncertainty representations [12], applies *de facto* for the confidence language currently used by the IPCC, the U.S. Defense Intelligence Agency and National Intelligence Council [38, 13], which involve verbal confidence qualifiers (e.g. *low*, *medium*, *high*). When two experts profess 'high confidence' in a particular probability assessment, how are they sure that they are talking about the same confidence level?

This can be thought of as a *calibration problem*: the challenge is to calibrate confidence levels across agents. Our proposed approach to this challenge is motivated by one method for calibrating (Bayesian) degrees of belief, which uses 'objective probabilities' or *chances*. When Ann and Bob say they have degree of belief 0.25 in an event F, we know they are expressing the same belief because they both agree that their degree of belief for the event is the same as their degree of belief for an event generated by 'objective random device', such as the event that the next ball drawn from an urn containing 5 red balls and 15 blue ones is red. Underlying this approach is the assumption that (different) rational agents agree on their degrees of belief for events for which they have been told the objective probabilities, or chances, and nothing else. Indeed, it is a consequence

of the Principal Principle [37] that, for any rational agent provided with the information that the chance of event E is x, and no further information relevant to this event, her degree of belief in E, $c(E) = x$.[13] Applying this with E being the event of drawing red from the aforementioned urn and $x = 0.25$, Ann and Bob both give the same degree of belief for E—$c_{Ann}(E) = c_{Bob}(E) = 0.25$—and hence, since they agree that this is the same as their degree of belief in F, we can conclude that they have the same degree of belief in F.

Note that this calibration of the probability language depends only on the *ordinal* properties of probabilities. To see this, consider Ann and Bob's qualitative probabilities $\succcurlyeq_{Ann}$ and $\succcurlyeq_{Bob}$ (in the sense of [43]). $E \succcurlyeq_{Ann} F$ if in Ann's subjective judgement, E is more likely than F. (In terms of betting behaviour, she prefers betting on E *ceteris paribus* to betting on F.) If these agents had real-valued credence functions, c_{Ann}, c_{Bob}, then for any pair of events E, F, $E \succcurlyeq_{Ann} F$ if and only if $c_{Ann}(E) \geq c_{Ann}(F)$. (Axioms on $\succcurlyeq_{Ann}, \succcurlyeq_{Bob}$ guarantee the existence of such functions; beyond transitivity, no such axioms need be imposed here.) Consider events E_x for $x \in X \subseteq [0, 1]$ with X closed, where all that is known about E_x is that its chance is x. For instance, they could be the events that red is drawn from urns with certain compositions. There is an objective 'more likely' order, $\succcurlyeq_l$ on such events, with $E_x \succcurlyeq_l E_y$ if and only if $x \geq y$. Note in particular that this relation is transitive and complete on $\{E_x\}_{x \in X}$. The map $\psi_{Ann} : \Sigma \to [0, 1]$, defined by $\psi_{Ann}(E) = \min\{x \in [0, 1] : E_x \succcurlyeq_{Ann} E\}$, assigns a value in X to each event, according to how it compares to the chance events E_x. It can be thought of as assigning 'degree-of-belief values' to events according to Ann's beliefs, on the scale represented by $\{E_x\}_{x \in X}$. A similar map can be defined for Bob. If these maps are to calibrate these agents' degrees of belief, then the map $\psi_{Bob}^{-1} \circ \psi_{Ann}$ needs to be order preserving: for all events E, F and all $E' \in \psi_{Bob}^{-1} \circ \psi_{Ann}(E)$ and $F' \in \psi_{Bob}^{-1} \circ \psi_{Ann}(F)$, $E' \succcurlyeq_{Bob} F'$ if and only if $E \succcurlyeq_{Ann} F$. In other words, more likely events according to one of the agents are assigned to a degree-of-belief level that is considered more likely according to the other. It is straightforward to see that:

Fact 1 *If, for all $x, y \in X$, $E_x \succcurlyeq_{Ann} E_y$ if and only if $E_x \succcurlyeq_{Bob} E_y$ if and only if $E_x \succcurlyeq_l E_y$, then $\psi_{Bob}^{-1} \circ \psi_{Ann}$ is order-preserving.*

The condition $E_x \succcurlyeq_{Ann} E_y$ if and only if $E_x \succcurlyeq_l E_y$ is just the Principal Principle mentioned above, applied to Ann's qualitative probability relation. So this result says that this

---

13. There has been debate on the Principal Principle in the philosophical literature, often focussing on its metaphysical implications, and purportedly turning on issues relating to its proper formulation [e.g. 21, 39]. We only require the special case of the principle applying to calibration devices here, which is generally uncontroversial: it avoids metaphysical subtleties and is implied by all reasonable existing versions of which we are aware. The same holds for the (new) weight-of-evidence version of the principle proposed below.

ordinal version of the Principal Principle is sufficient for the chance events $\{E_x\}_{x \in X}$ to calibrate Ann's and Bob's degree-of-belief levels. Whilst taking richer scales ($X \to [0,1]$) and assuming standard continuity properties of the qualitative probability relations will yield a full $[0,1]$-map for each agent—corresponding to his or her credence function—for communication and elicitation purposes, a coarser scale is usually sufficient and maximally feasible. This calibration argument has been conducted entirely at the ordinal level, and does not require the additivity or normalisation properties of probability measures.

The same logic can be applied to confidence, relying on the relation with the weight of evidence. Recall that, under the separation of credal judgements and confidence in them under this model, the former track aspects of the 'balance of evidence'—in the typical examples of draws from urns, the frequency of a given colour—whereas the latter tracks the weight—often associated to the size of the sample. After observing 100 draws, 50% of which were red, one may hold the credal judgement $p(Red) \geq 0.4$ with a fairly high level of confidence; if 50% out of two draws were red, one's confidence would reasonably be much lower. This suggests that confidence could be calibrated using a range of situations where the 'objective' weight of evidence—the number of draws in the example—varies.

More precisely, consider events E_n, $n \in \mathbb{N}$, with each E_n being the event that the next ball drawn from a given urn is red, where the only information possessed about the urn is: (a) it contains 20 balls, each of which is red or blue; (b) $2n$ draws from the urn have been observed (with replacement), precisely n of which were red. Let P_n be the credal statement that $p(E_n) \geq 0.4$ ($P_n = \{p \in \Delta : p(E_n) \geq 0.4\}$). Between P_n and $P_{n'}$ with $n > n'$, the credal statement and the frequency of red remain fixed; only the weight of evidence changes, being larger for n. In the same way that it is an objective fact that the chance of picking red from an urn with 10 red balls out of 20 is larger than the chance of picking red from an urn with 5 red balls out of 20, it seems to be an objective fact that the weight of evidence for P_{50} is greater than the weight of evidence for P_5. Using $\succeq_{w.o.e}$ for the '(objectively) has more weight of evidence' relation, we have that $P_n \succeq_{w.o.e.} P_{n'}$ whenever $n \geq n'$.

Now let Ξ_{Ann} and Ξ_{Bob} be Ann and Bob's confidence rankings, and define the map $\gamma_{Ann} : \Xi_{Ann} \to \mathbb{N}$ by $\gamma_{Ann}(\mathscr{C}) = \min\{n : \mathscr{C} \subseteq P_n\}$, and similarly for γ_{Bob}. Confidence levels are characterised by the *lowest* n for which the agent holds the belief P_n with the level of confidence in question. This can be thought of as assigning 'confidence values' on a numerical scale. As for the probability case above, to be an appropriate calibration, the map $\gamma_{Bob}^{-1} \circ \gamma_{Ann}$ relating Ann's and Bob's confidence levels needs to be order-preserving.[14]

---

14. For all $\mathscr{C}, \mathscr{D} \in \Xi_{Ann}$, $\mathscr{C}' \in \gamma_{Bob}^{-1} \circ \gamma_{Ann}(\mathscr{C})$ and $\mathscr{D}' \in \gamma_{Bob}^{-1} \circ \gamma_{Ann}(\mathscr{D})$, $\mathscr{C}' \subseteq \mathscr{D}' \Leftrightarrow \mathscr{C} \subseteq \mathscr{D}$.

Again, we have a condition under which this is the case (where $\succeq_\Xi$ is as defined in Section 2.1).

Fact 2 *If, for all $n, n' \in \mathbb{N}$, $P_n \succeq_{\Xi_{Ann}} P_{n'}$ if and only if $P_n \succeq_{\Xi_{Bob}} P_{n'}$ if and only if $P_n \succeq_{w.o.e.} P_{n'}$, then $\gamma_{Bob}^{-1} \circ \gamma_{Ann}$ is order-preserving.*

The condition $P_n \succeq_{\Xi_{Ann}} P_{n'}$ if and only if $P_n \succeq_{w.o.e.} P_{n'}$ can be thought as a *weight-of-evidence Principal Principle*: if the only information available supports one credal statement objectively with more weight of evidence than another, then rational agents hold the former with more confidence. As for the Bayesian case, refining the 'scale' and under continuous confidence rankings [24], one can in principle obtain a finer distinction in confidence levels; again, for communication and elicitation practice, there is little indication that further refinements would be motivated.

The map γ thus defines a numerical scale for confidence, which, under a weight-of-evidence version of the Principal Principle, calibrates confidence across agents. In particular, it permits interpersonal comparison of confidence levels. Ann and Bob can be said to be professing the same confidence in a given judgement, if each of them judge the confidence to be the same as the confidence in P_n—in other words, if each of them associate it to the same n on the scale. Such scales thus resolve the communication problem, in theory. In fact, confidence can be elicited on such scale—protocols can be developed allowing one to 'extract' agents' confidence in beliefs on such a scale, using their confidence comparisons to the scale judgements P_n. A web tool permitting elicitation on such a scale has been developed, and is available here: http://confidence.hec.fr/app/.

We close with some remarks about the choice of scale. Consider two 'calibration sets' of credal statements $\{P_n\}_{n \in \mathbb{N}}$ and $\{P'_n\}_{n \in \mathbb{N}}$, with the objective weight of evidence increasing in n in both cases. What is the relationship between them? Recall that, for a set of credal statements to serve as the basis of a scale, the objective weight of evidence relation $\succeq_{w.o.e.}$ has to be complete and transitive on it. Suppose that $\succeq_{w.o.e.}$ is transitive and complete on $\{P_n\}_{n \in \mathbb{N}} \cup \{P'_n\}_{n \in \mathbb{N}}$, so, in particular, it can compare each P'_n on the $\{P_n\}_{n \in \mathbb{N}}$ scale. Then $\{P_n\}_{n \in \mathbb{N}} \cup \{P'_n\}_{n \in \mathbb{N}}$ (or any superset on which $\succeq_{w.o.e.}$ is transitive and complete) calibrates confidence; the previous numerical scales can thus be compared by translating them into a finer scale. This is well-known for the credence case: draws from urns with 20 balls and rolls of a fair die each calibrate degrees of belief (up to the nearest $\frac{1}{20}$ and $\frac{1}{6}$): although events in one do not directly correspond to those in another, what counts is that they can be compared in terms of chance, and 'embedded' into a richer common scale.

So what really counts for comparing different scales is whether $\succeq_{w.o.e.}$ is transitive and complete over the union. Is this the case for any pair of 'calibration sets' of credal statements $\{P_n\}_{n \in \mathbb{N}}$ and $\{P'_n\}_{n \in \mathbb{N}}$? Note that the answer may be

no even for degrees of belief. If each event in $\{E_x\}_{x \in X}$ or $\{E'_x\}_{x \in X'}$ has a precise objective chance, then $\succcurlyeq_l$ is complete and transitive over the union of these families. However, this would not be universally the case if, as argued by [20], there are events with indeterminate (yet objective) chances. The order $\succcurlyeq_l$ is complete and transitive over each of the families $\{E_0, E_{[0.15,0.35]}, E_{[0.4,0.6]}, E_{[0.65,0.85]}, E_1\}$ and $\{E'_0, E'_{0.2}, E'_{0.4}, E'_{0.6}, E'_{0.8}, E'_1\}$, but it does not order $E_{[0.65,0.85]}$ (i.e. the event with indeterminate chance $[0.65, 0.85]$) and $E'_{0.8}$, so it is not complete over the union of these families. Calibrating degrees of belief on chances *cannot and does not use* any scale of chance events, but restricts to a subclass of scales which are comparable in the previous sense: the scales containing only events with precise chances.

Similarly, if not all scales $\{P_n\}_{n \in \mathbb{N}}$ and $\{P'_n\}_{n \in \mathbb{N}}$ were comparable (in the sense of their union being completely ordered by $\succeq_{w.o.e.}$), one could restrict attention to a maximally comparable scale. In the end, the issue of comparability will turn on properties of the objective weight of evidence relation. For instance, under an account according to which the weight of evidence for P_n is the posterior Bayesian probability of P_n after conditionalisation of a uniform prior on the information received in the E_n situation, $\succeq_{w.o.e.}$ is complete across any union of such scales, and all such scales are comparable. The question of objective weight of evidence, and hence of comparability, goes beyond the scope of this paper, and must thus be left for future research.

4. Conclusion

In this paper, we have considered how a recently-defended approach to belief and decision, which gives pride of place to the notion of confidence in beliefs, can render the Keynesian notion of weight of evidence and mobilise it in uncertainty reporting. It seems to be able to capture the distinction between weight and other aspects of evidence, via the difference between confidence in a credal judgement and the credal judgement at issue. It does not require that weight and balance—confidence and credal precision—be traded off into a single credal set. When connected to decision, this supports a clear separation of the belief and value factors of the model: an important quality in contexts where these factors are the responsibility of different agents. Finally, we consider the communication of confidence, and propose a resolution to the problem of interpersonal confidence calibration, evoking a weight-of-evidence version of the Principal Principle.

Acknowledgments

References

[1] Mohammed Abdellaoui, Brian Hill, Emmanuel Kemel, and Hela Maafi. The Evolution of Ambiguity Attitudes through Learning. mimeo, HEC Paris, 2016.

[2] Kenneth J Arrow. *Essays in the Theory of Risk Bearing*. Markham Publishing., Chicago, 1971.

[3] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 1985. ISBN 978-1-4757-4286-2.

[4] Truman F Bewley. Knightian decision theory. Part I. *Decisions in Economics and Finance*, 25(2):79–110, 2002.

[5] Richard Bradley, Casey Helgeson, and Brian Hill. Climate Change Assessments: Confidence, Probability, and Decision. *Philosophy of Science*, 84(3):500–522, 2017.

[6] David V. Budescu and Thomas S. Wallsten. Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation*, 32:275–318, 1995.

[7] David V. Budescu, Han-Hui Por, and Stephen B. Broomell. Effective communication of uncertainty in the IPCC reports. *Climatic change*, 113(2):181–200, 2012. doi:10.1007/s10584-011-0330-3.

[8] David V. Budescu, Han-Hui Por, Stephen B. Broomell, and Michael Smithson. The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, 4(6):508–512, 2014. doi:10.1038/nclimate2194.

[9] Roger M. Cooke. *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press, 1991.

[10] Heather Douglas. *Science, policy, and the value-free ideal*. University of Pittsburgh Press, 2009.

[11] D Ellsberg. Risk, Ambiguity, and the Savage Axioms. *Quart. J. Econ.*, 75:643–669, 1961.

[12] Scott Ferson, Cliff A. Joslyn, Jon C. Helton, William L. Oberkampf, and Kari Sentz. Summary from the epistemic uncertainty workshop: consensus amid diversity. *Reliability Engineering & System*

Safety, 85(1–3):355–369, July 2004. ISSN 0951-8320. doi:10.1016/j.ress.2004.03.023.

[13] Jeffrey A. Friedman and Richard Zeckhauser. Analytic Confidence and Political Decision-Making: Theoretical Principles and Experimental Evidence From National Security Professionals. *Political Psychology*, 39(5):1069–1087, 2018.

[14] Peter Gärdenfors and Nils-Eric Sahlin. Unreliable probabilities, risk taking, and decision making. *Synthese*, 53(3):361–386, 1982. doi:10.1007/BF00486156.

[15] Paolo Ghirardato and Massimo Marinacci. Ambiguity Made Precise: A Comparative Foundation. *J. Econ. Theory*, 102(2):251–289, 2002. ISSN 0022-0531. doi:10.1006/jeth.2001.2815. URL http://www.sciencedirect.com/science/article/B6WJ3-4576D74-1/2/f3d013b3d11d4e1500d6f0805a6f507b.

[16] Ithzak Gilboa and Massimo Marinacci. Ambiguity and the Bayesian Paradigm. In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*. 2013.

[17] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *J. Math. Econ.*, 18(2):141–153, 1989. doi:10.1016/0304-4068(89)90018-9.

[18] Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler. Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, 25(3):285–296, nov 2009. ISSN 0266-2671. doi:10.1017/S0266267109990241.

[19] Irving John Good. *Probability and the Weighing of Evidence*. Griffin, London, 1950.

[20] Alan Hájek and Michael Smithson. Rationality and indeterminate probabilities. *Synthese*, 187(1):33–48, July 2012. ISSN 1573-0964. doi:10.1007/s11229-011-0033-3.

[21] Ned Hall. Correcting The Guide to Objective Chance. *Mind*, 103(412):505–517, 1994. ISSN 0026-4423. URL http://www.jstor.org/stable/2254398.

[22] Casey Helgeson, Richard Bradley, and Brian Hill. Combining probability with qualitative degree-of-certainty metrics in assessment. *Climatic Change*, 149(3):517–525, August 2018. ISSN 1573-1480. doi:10.1007/s10584-018-2247-6.

[23] Brian Hill. Confidence in preferences. *Social Choice and Welfare*, 39(2):273–302, 2012.

[24] Brian Hill. Confidence and decision. *Games and Economic Behavior*, 82:675–692, 2013. ISSN 0899-8256. doi:10.1016/j.geb.2013.09.009.

[25] Brian Hill. Confidence as a Source of Deferral. HEC Paris Research Paper No. ECO/SCD-2014-1060, 2015.

[26] Brian Hill. Incomplete preferences and confidence. *Journal of Mathematical Economics*, 65:83–103, August 2016. ISSN 0304-4068. doi:10.1016/j.jmateco.2016.05.007. URL http://www.sciencedirect.com/science/article/pii/S0304406816300301.

[27] Brian Hill. Confidence in Beliefs and Rational Decision Making. *Economics & Philosophy*, 35(2): 223–258, July 2019. ISSN 0266-2671, 1474-0028. doi:10.1017/S0266267118000214.

[28] Brian Hill. Updating Confidence in Beliefs. HEC Paris Research Paper, 2019.

[29] IPCC. Summary for policymakers. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[30] James M Joyce. How Probabilities Reflect Evidence. *Philosophical Perspectives*, 2005. doi:10.1111/j.1520-8583.2005.00058.x.

[31] James M Joyce. A Defense of Imprecise Credences in Inference and Decision Making. In Tamar Szabo Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology*, volume 4. OUP, 2011.

[32] John Maynard Keynes. *A Treatise on Probability*. Dover Publications, 1921.

[33] Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji. A Smooth Model of Decision Making under Ambiguity. *Econometrica*, 73(6):1849–1892, 2005. doi:10.1111/j.1468-0262.2005.00640.x.

[34] Isaac Levi. On Indeterminate Probabilities. *Journal of Philosophy*, 71:391–418, 1974.

[35] Isaac Levi. *Hard Choices. Decision making under unresolved conflict*. Cambridge University Press, Cambridge, 1986.

[36] Isaac Levi. The Weight of Argument. In Silva Marzetti Dall'Aste Brandolini and Roberto Scazzieri, editors, *Fundamental Uncertainty: Rationality and Plausible Reasoning*, pages 39–58. Palgrave Macmillan UK, London, 2011. ISBN 978-0-230-30568-7. doi:10.1057/9780230305687_3.

[37] David K Lewis. A Subjectivist's Guide to Objective Chance. In *Philosophical Papers*, volume 2, pages 83–132. OUP, Oxford, 1986.

[38] Michael D. Mastrandrea, Christopher B. Field, Thomas F. Stocker, Ottmar Edenhofer, Kristie L. Ebi, David J. Frame, Hermann Held, Elmar Kriegler, Katharine J. Mach, and Patrick R. Matschoss. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. *Intergovernmental Panel on Climate Change (IPCC)*, 2010. URL http://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf.

[39] Christopher JG Meacham. Two mistakes regarding the principal principle. *the British Journal for the Philosophy of Science*, 61(2):407–431, 2009.

[40] M. Granger Morgan. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, 111(20):7176–7184, 2014. doi:10.1073/pnas.1319946111.

[41] Robert F Nau. Indeterminate Probabilities on Finite Sets. *The Annals of Statistics*, 20(4):1737–1767, 1992. ISSN 00905364. URL http://www.jstor.org/stable/2242366.

[42] John W Pratt. Risk Aversion in the Small and in the Large. *Econometrica*, 32:122–136, 1964.

[43] Leonard J Savage. *The Foundations of Statistics*. Dover, New York, 1954.

[44] George Lennox Sharman Shackle. *Expectation in economics*. Cambridge University Press, 1949.

[45] Brian Skyrms. Resiliency, Propensities, and Causal Necessity. *The Journal of Philosophy*, 74(11):704–713, November 1977. doi:10.2307/2025774.

[46] Michael Smithson. Elicitation. In Thomas Augustin, Frank PA Coolen, Gert de Cooman, and Matthias CM Troffaes, editors, *Introduction to imprecise probabilities*, pages 318–328. John Wiley & Sons, 2014. doi:10.1002/9781118763117.ch15.

[47] Katie Steele. The scientist qua policy advisor makes value judgments. *Philosophy of Science*, 79(5):893–904, 2012. URL http://www.jstor.org/stable/10.1086/667842.

[48] Erica Thompson, Roman Frigg, and Casey Helgeson. Expert Judgment for Climate Change Adaptation. *Philosophy of Science*, 83(5):1110–1121, June 2016. ISSN 0031-8248. doi:10.1086/687942.

[49] Peter Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

[50] Brian Weatherson. Keynes, uncertainty and interest rates. *Cambridge Journal of Economics*, 26(1):47–62, 2002. ISSN 0309-166X. URL https://www.jstor.org/stable/23600357.

[51] Ilan Yaniv and Dean P. Foster. Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124(4):424, 1995. doi:10.1037/0096-3445.124.4.424.

[52] Ilan Yaniv and Dean P. Foster. Precision and accuracy of judgmental estimation. *Journal of behavioral decision making*, 10(1):21–32, 1997. doi:10.1002/(SICI)1099-0771(199703)10:1<21::AID-BDM243>3.0.CO;2-G.