Imprecise Probabilities

First published Sat Dec 20, 2014; substantive revision Tue Feb 19, 2019

It has been argued that imprecise probabilities are a natural and intuitive way of overcoming some of the issues with orthodox precise probabilities. Models of this type have a long pedigree, and interest in such models has been growing in recent years. This article introduces the theory of imprecise probabilities, discusses the motivations for their use and their possible advantages over the standard precise model. It then discusses some philosophical issues raised by this model. There is also a historical appendix which provides an overview of some important thinkers who appear sympathetic to imprecise probabilities.

1. Introduction
   1.1 A summary of terminology
   1.2 Some important distinctions

2. Motivations
   2.1 Ellsberg decisions
   2.2 Incompleteness and incomparability
   2.3 Weight of evidence, balance of evidence
   2.4 Suspending judgement
   2.5 Unknown correlations
   2.6 Nonprobabilistic chances
   2.7 Group belief

3. Philosophical questions for IP
   3.1 Dilation
   3.2 Belief inertia
   3.3 Decision making
   3.4 Interpreting IP
      3.4.1 What is a belief?
      3.4.2 What is a belief in \( X \)?
3.5 Regress
3.6 What makes a good imprecise belief?
4. Summary
Bibliography
Academic Tools
Other Internet Resources
Related Entries

1. Introduction

Probability theory has been a remarkably fruitful theory, with applications in almost every branch of science. In philosophy, some important applications of probability theory go by the name Bayesianism; this has been an extremely successful program (see for example Howson and Urbach 2006; Bovens and Hartmann 2003; Talbott 2008). But probability theory seems to impute much richer and more determinate attitudes than seems warranted. What should your rational degree of belief be that global mean surface temperature will have risen by more than four degrees by 2080? Perhaps it should be 0.75? Why not 0.75001? Why not 0.7497? Is that event more or less likely than getting at least one head on two tosses of a fair coin? It seems there are many events about which we can (or perhaps should) take less precise attitudes than orthodox probability requires. Among the reasons to question the orthodoxy, it seems that the insistence that states of belief be represented by a single real-valued probability function is quite an unrealistic idealisation, and one that brings with it some rather awkward consequences that we shall discuss later. Indeed, it has long been recognised that probability theory offers only a rather idealised model of belief. As far back as the mid-nineteenth century, we find George Boole saying:

"It would be unphilosophical to affirm that the strength of that expectation, viewed as an emotion of the mind, is capable of being referred to any numerical standard. (Boole 1958 [1854]: 244)"

For these, and many other reasons, there is growing interest in Imprecise Probability (IP) models. Broadly construed, these are models of belief that go beyond the probabilistic orthodoxy in one way or another.

IP models are used in a number of fields including:

- Statistics (Walley 1991; Ruggeri et al. 2005; Augustin et al. 2014)
- Psychology of reasoning (Pfeifer and Kleiter 2007)
- Linguistic processing of uncertainty (Wallsten and Budescu 1995)
- Neurological response to ambiguity and conflict (Smithson and Pushkarskaya 2015)
- Philosophy (Levi 1980; Joyce 2011; Sturgeon 2008; Kaplan 1983; Kyburg 1983)
- Behavioural economics (Ellsberg 1961; Camerer and Weber 1992; Smithson and Campbell 2009)
- Mathematical economics (Gilboa 1987)
- Engineering (Ferson and Ginzburg 1996; Ferson and Hajagos 2004; Oberguggenberger 2014)
- Computer science (Cozman 2000; Cozman and Walley 2005)
- Scientific computing (Oberkampf and Roy 2010, chapter 13)
- Physics (Suppes and Zanotti 1991; Hartmann and Suppes 2010; Frigg et al. 2014)

This article identifies a variety of motivations for IP models; introduces various formal models that are broadly in this area; and discusses some open problems for these frameworks. The focus will be on formal models of belief.

1.1 A summary of terminology
Throughout the article I adopt the convention of discussing the beliefs of an arbitrary intentional agent whom I shall call “you”. Prominent advocates of IP (including Good and Walley) adopt this convention.

This article is about formal models of belief and as such, there needs to be a certain amount of formal machinery introduced. There is a set of states \( \Omega \) which represents the ways the world could be. Sometimes \( \Omega \) is described as the set of “possible worlds”. The objects of belief—the things you have beliefs about—can be represented by subsets of the set of ways the world could be \( \Omega \). We can identify a proposition \( X \) with the set of states which make it true, or, with the set of possible worlds where it is true. If you have beliefs about \( X \) and \( Y \) then you also have beliefs about “\( X \cap Y \)”, “\( X \cup Y \)” and “\( \neg X \)” “\( X \) and \( Y \)”, “\( X \) or \( Y \)”; and it is not the case that \( X \)” respectively. The set of objects of belief is the power set of \( \Omega \), or if \( \Omega \) is infinite, some measurable algebra of the subsets of \( \Omega \).

The standard view of degree of belief is that degrees of belief are represented by real numbers and belief states by probability functions; this is a normative requirement. Probability functions are functions, \( p \), from the algebra of beliefs to real numbers satisfying:

- \( 0 = p(\emptyset) \leq p(X) \leq p(\Omega) = 1 \)
- If \( X \cap Y = \emptyset \) then \( p(X \cup Y) = p(X) + p(Y) \)

So if your belief state or doxastic state is represented by \( p \), then your degree of belief in \( X \) is the value assigned to \( X \) by \( p \); that is, \( p(X) \).

Further, learning in the Bayesian model of belief is effected by conditionalisation. If you learn a proposition \( E \) (and nothing further) then your post-learning belief in \( X \) is given by \( p(X \mid E) = p(X \cap E)/p(E) \).

The alternative approach that will be the main focus of this article is the approach that represents belief by a set of probability functions instead of a single probability. So instead of having some \( p \) represent your belief state, you have \( P \), a set of such functions. van Fraassen (1990) calls this your representer, Levi calls it a credal set. I will discuss various ways you might interpret the representer later but for now we can think of it as follows. Your representer is a credal committee: each probability function in it represents the opinions of one member of a committee that, collectively, represents your beliefs.

From these concepts we can define some “summary functions” that are often used in discussions of imprecise probabilities. Often, it is assumed that your degree of belief in a proposition, \( X \), is represented by \( P(X) = \{ p(X) : p \in P \} \). I will adopt this notational convention, with the proviso that I don’t take \( P(X) \) to be an adequate representation of your degree of belief in \( X \). Your lower envelope of \( X \) is: \( \underline{P}(X) = \inf P(X) \). Likewise, your upper envelope is \( \overline{P}(X) = \sup P(X) \). They are conjugates of each other in the following sense: \( \overline{P}(X) = 1 - \underline{P}(\neg X) \).

The standard assumption about updating for sets of probabilities is that your degree of belief in \( X \) after learning \( E \) is given by \( P(X \mid E) = \{ p(X \mid E), p \in P, p(E) > 0 \} \). Your belief state after having learned \( E \) is \( P(\cdot \mid E) = \{ p(\cdot \mid E), p \in P, p(E) > 0 \} \). That is, by the set of conditional probabilities.

I would like to emphasise already that these summary functions—\( P(\cdot) \), \( \underline{P}(\cdot) \) and \( \overline{P}(\cdot) \)—are not properly representative of your belief. Information is missing from the picture. This issue will be important later, in our discussion of dilation.

We shall need to talk about decision making so we shall introduce a simple model of decisions in terms of gambles. We can view bounded real valued functions \( f \) as “gambles” that are functions from some set \( \Omega \) to real numbers. A gamble \( f \) pays out \( f(\omega) \) if \( \omega \) is the true state. We assume that
There are a number of distinctions that it is important to make in what follows.

An important parameter in an IP theory is the normative force the theory is supposed to have. Is imprecision obligatory or is it merely permissible? Is it always permissible/obligatory, or only sometimes? Or we might be interested in a purely descriptive project of characterising the credal states of actual agents, with no interest in normative questions. This last possibility will concern us little in this article.

It is also helpful to distinguish belief itself from the elicitation of that belief and also from your introspective access to those beliefs. The same goes for other attitudes (values, utilities and so on). It may be that you have beliefs that are not amenable to (precise) elicitation, in practice or even in principle. Likewise, your introspective access to your own beliefs might be imperfect. Such imperfections could be a source of imprecision. Bradley (2009) distinguishes many distinct sources of imperfect introspection. The imperfection could arise from your unawareness of the prospect in question, the boundedness of your reasoning, ignorance of relevant contingencies, or because of conflict in your evidence or in your values (pp. 240–241). See Bradley and Drechsler (2014) for further discussion of types of uncertainty.

There are a variety of aspects of a body of evidence that could make a difference to how you ought to respond to it. We can ask how much evidence there is (weight of evidence). We can ask whether the evidence is balanced or whether it tells heavily in favour of one hypothesis over another (balance of evidence). Evidence can be balanced because it is incomplete: there simply isn’t enough of it. Evidence can also be balanced if it is conflicted: different pieces of evidence favour different hypotheses. We can further ask whether evidence tells us something specific—like that the bias of a coin is 2/3 in favour of heads—or unspecific—like that the...
bias of a coin is between 2/3 and 1 in favour of heads. This specificity should be distinguished from vagueness or indeterminacy of evidence: that a coin has bias about 2/3 is vague but specific, while that a coin has bias definitely somewhere between 2/3 and 1 is determinate but unspecific. Likewise, a credal state could be indeterminate, fuzzy, or it could be unspecific, or it could be both. It seems like determinate but unspecific belief states will be rarer than indeterminate ones.

Isaac Levi (1974, 1985) makes a distinction between “imprecise” credences and “indeterminate” credences (the scare quotes are indicating that these aren’t uses of the terms “imprecise” and “indeterminate” that accord with the usage I adopt in this article). The idea is that there are two distinct kinds of belief state that might require a move to an IP representation of belief. An “imprecise” belief in Levi’s terminology is an imperfectly introspected or elicited belief in mine, while an “indeterminate” belief is a (possibly) perfectly introspected belief that is still indeterminate or unspecific (or both). Levi argues that the interesting phenomenon is “indeterminate” credence. Walley (1991) also emphasises the distinction between cases where there is a “correct” but unknown probability and cases of “indeterminacy”.

There is a further question about the interpretation of IP that cross-cuts the above. This is the question of whether we understand $P$ as a “complete” or “exhaustive” representation of your beliefs, or whether we take the representation to be incomplete or non-exhaustive. Let’s talk in terms of the betting interpretation for a moment. The exhaustive/non-exhaustive distinction can be drawn by asking the following question: does $P$ capture all and only your dispositions to bet or does $P$ only partially capture your dispositions to bet? Walley emphasises this distinction and suggests that most models are non-exhaustive.

Partly because of Levi’s injunction to distinguish “imprecise” from “indeterminate” belief, some have objected to the use of the term “imprecise probability”. Using the above distinction between indeterminate, unspecific and imperfectly introspected belief, we can keep separate the categories Levi wanted to keep separate all without using the term “imprecise”. We can then use “imprecise” as an umbrella term to cover all these cases of lack of precision. Conveniently, this allows us to stay in line with the wealth of formal work on “Imprecise Probabilities” which term is used to cover cases of indeterminacy. This usage goes back at least to Peter Walley’s influential book *Statistical Reasoning with Imprecise Probabilities* (Walley 1991).

So, “Imprecise” is not quite right, but neither is “Probability” since the formal theory of IP is really about *previsions* (sort of expectations) rather than just about probability (expectations of indicator functions). Helpfully, if I abbreviate Imprecise Probability to “IP” then I can exploit some useful ambiguities.

### 2. Motivations

Let’s consider, in general terms, what sort of motivations one might have for adopting models that fall under the umbrella of IP. The focus will be on models of rational belief, since these are the models that philosophers typically focus on, although it is worth noting that statistical work using IP isn’t restricted to this interpretation. Note that no one author endorses all of these arguments, and indeed, some authors who are sympathetic to IP have explicitly stated that they don’t consider certain of these arguments to be good (for example Mark Kaplan does not endorse the claim that concerns about descriptive realism suggest allowing incompleteness).

#### 2.1 Ellsberg decisions

Partly because of Levi’s injunction to distinguish “imprecise” from “indeterminate” belief, some have objected to the use of the term “imprecise probability”. Using the above distinction between indeterminate, unspecific and imperfectly introspected belief, we can keep separate the categories Levi wanted to keep separate all without using the term “imprecise”. We can then use “imprecise” as an umbrella term to cover all these cases of lack of precision. Conveniently, this allows us to stay in line with the wealth of formal work on “Imprecise Probabilities” which term is used to cover cases of indeterminacy. This usage goes back at least to Peter Walley’s influential book *Statistical Reasoning with Imprecise Probabilities* (Walley 1991).

So, “Imprecise” is not quite right, but neither is “Probability” since the formal theory of IP is really about *previsions* (sort of expectations) rather than just about probability (expectations of indicator functions). Helpfully, if I abbreviate Imprecise Probability to “IP” then I can exploit some useful ambiguities.

### 2. Motivations

Let’s consider, in general terms, what sort of motivations one might have for adopting models that fall under the umbrella of IP. The focus will be on models of rational belief, since these are the models that philosophers typically focus on, although it is worth noting that statistical work using IP isn’t restricted to this interpretation. Note that no one author endorses all of these arguments, and indeed, some authors who are sympathetic to IP have explicitly stated that they don’t consider certain of these arguments to be good (for example Mark Kaplan does not endorse the claim that concerns about descriptive realism suggest allowing incompleteness).

#### 2.1 Ellsberg decisions
There are a number of examples of decision problems where we are intuitively drawn to go against the prescriptions of precise probabilism. And indeed, many experimental subjects do seem to express preferences that violate the axioms. IP offers a way of representing these intuitively plausible and experimentally observed choices as rational. One classic example of this is the Ellsberg problem (Ellsberg 1961).

I have an urn that contains ninety marbles. Thirty marbles are red. The remainder are blue or yellow in some unknown proportion.

Consider the indicator gambles for various events in this scenario. Consider a choice between a bet that wins if the marble drawn is red (I), versus a bet that wins if the marble drawn is blue (II). You might prefer I to II since I involves risk while II involves ambiguity. A prospect is risky if its outcome is uncertain but its outcomes occur with known probability. A prospect is ambiguous if the outcomes occur with unknown or only partially known probabilities. Now consider a choice between a bet that wins if the marble drawn is not blue (III) versus a bet that wins if the marble drawn is not red (IV). Now it is III that is ambiguous, while IV is unambiguous but risky, and thus IV might seem better to you if you preferred risky to ambiguous prospects. Such a pattern of preferences (I preferred to II but IV preferred to III) cannot be rationalised as the choices of a precise expected utility maximiser. The gambles are summarised in the table.

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>B</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: The Ellsberg bets. The urn contains 30 red marbles and 60 blue/yellow marbles.

Let the probabilities for red, blue and yellow marbles be \( r, b \) and \( y \) respectively. If you were an expected utility maximiser and preferred I to II, then \( r > b \) and a preference for IV over III entails that \( r + y < y + b \). No numbers can jointly satisfy these two constraints. Therefore, no probability function is such that an expected utility maximiser with that probability would choose in the way described above. While by no means universal, these preferences are a robust feature of many experimental subjects’ response to this sort of example (Camerer and Weber 1992; Fox and Tversky 1995). Some experiments suggest that Ellsberg-type patterns of preference are rarer than normally recognised (Binmore et al. 2012; Voorhoeve et al. 2016). For more on ambiguity attitudes, see Trautmann and van der Kuilen (2016).

The imprecise probabilist can model the situation as follows: \( P(R) = 1/3, P(B) = P(Y) = [0, 2/3] \). Note that this expression of the belief state misses out some important details. For example, for all \( p \in P \), we have \( p(B) = 2/3 - p(Y) \). For the point being made here, this detail is not important. Modelling the ambiguity allows us to rationalise real agents’ preferences for bets on red. To flesh this story out would require a lot more to be said about decision making, (see section 3.3) but the intuition is that aversion to ambiguity explains the preference for I over II and IV over III.

As Steele (2007) points out, the above analysis rationalises the Ellsberg choices only if we are dealing with genuinely indeterminate or unspecific beliefs. If we were dealing with a case of imperfectly introspected belief then there would exist some \( p \) in the representor such that rational choices maximise \( E_p \). For the Ellsberg choices, there is no such \( p \).

This view on the lessons of the Ellsberg game is not uncontroversial. Al-Najjar and Weinstein (2009) offer an alternative view on the interpretation of the Ellsberg preferences. Their view is that the distinctive pattern of
Ellsberg choices is due to agents applying certain heuristics to solve the decisions that assume that the odds are manipulable. In real-life situations, if someone offers you a bet, you might think that they must have some advantage over you in order for it to be worth their while offering you the bet. Such scepticism, appropriately modelled, can yield the Ellsberg choices within a simple game theoretic precise probabilistic model.

2.2 Incompleteness and incomparability

Various arguments for (precise) probabilism assume that some relation or other is complete. Whether this is a preference over acts, or some “qualitative probability ordering”, the relation is assumed to hold one way or the other between any two elements of the domain. This hardly seems like it should be a principle of rationality, especially in cases of severe uncertainty. That is—to take the preference example—it is reasonable to have no preference in either direction. This is an importantly different attitude to being indifferent between the options. Mark Kaplan argues this point as follows:

Both when you are indifferent between A and B and when you are undecided between A and B you can be said not to prefer either state of affairs to the other. Nonetheless, indifference and indecision are distinct. When you are indifferent between A and B, your failure to prefer one to the other is born of a determination that they are equally preferable. When you are undecided, your failure to prefer one to the other is born of no such determination. (Kaplan 1996: 5)

There is a standard behaviourist response to the claim that incomparability and indifference should be distinguished. In short, the claim is that it is a distinction that cannot be inferred from actual agents’ choice behaviour. Ultimately, in a given choice situation you must choose one of the options.

Which you choose can be interpreted as being (weakly) preferred. Joyce offers the following criticism of this appeal to behaviourism.

There are just too many things worth saying that cannot be said within the confines of strict behaviourism… The basic difficulty here is that it is impossible to distinguish contexts in which an agent’s behavior really does reveal what she wants from contexts in which it does not without appealing to additional facts about her mental state… An even more serious shortcoming is behaviourism’s inability to make sense of rationalizing explanations of choice behavior. (Joyce 1999: 21)

On top of this, behaviourists cannot make sense of the fact that incomparable goods are insensitive to small improvements. That is, if A and B are two goods that you have no preference between (for example, bets on propositions with imprecise probabilities) and if A+ is a good slightly better than A, then it might still be incomparable with B. This distinguishes incomparability from indifference, since indifference “ties” will be broken by small improvements. So the claim that there is no behavioural difference between indifference and incomparability is false.

Kaplan argues that not only is violating the completeness axiom permissible, it is, in fact, sometimes obligatory.

[My reason for rejecting as falsely precise the demand that you adopt a … set of preferences [that satisfy the preference axioms] is not the usual one. It is not that this demand is not humanly satisfiable. For if that were all that was wrong, the demand might still play a useful role as a regulative ideal—an ideal which might then be legitimately invoked to get you to “solve” your decision problem as the orthodox Bayesian would have you do. My complaint about the orthodox Bayesian demand is rather that it
He notes that it is not the case that it is always unreasonable or impossible for you to have precise beliefs: in that case precision could serve as a regulative ideal. Precise probabilism does still serve as something of a regulative ideal, but it is the belief of an ideal agent in an idealised evidential position. Idealised evidential positions are approximated by cases where you have a coin of a known bias. Precise probabilists and advocates of IP both agree that precise probabilism is an idealisation, and a regulative ideal. However, they differ as to what kind of idealisation is involved. Precise probabilists think that what precludes us from having precise probabilistic beliefs is merely a lack of computational power and introspective capacity. Imprecise probabilists think that even agents ideal in this sense might (and possibly should) fail to have precise probabilistic beliefs when they are not in an ideal evidential position.

At least some of the axioms of preference are not normative constraints. We can now ask what can be proved in the absence of the "purely structural"—non-normative—axioms? This surely gives us a handle on what is really required of the structure of belief.

It seems permissible to fail to have a preference between two options. Or it seems reasonable to fail to consider either of two possibilities more likely than the other. And these failures to assent to certain judgements is not the same as considering the two elements under consideration to be on a par in any substantive sense. That said, precise probabilism is serving as a regulative ideal. That is, precision might still be an unattained (possibly unattainable) goal that informs agents as to how they might improve their credences. Completeness of preference is what the thoroughly informed agent ought to have. Without complete preference, standard representation theorems don’t work. However, for each completion of the incomplete preference ordering—for each complete ordering that extends the incomplete preference relation—the theorem follows. So if we consider the set of probability functions that are such that some completion of the incomplete preference is represented by that function, then we can consider this set to be representing the beliefs associated with the incomplete preference. We also get, for each completion, a utility function unique up to linear transformation. This, in essence, was Kaplan’s position (see Kaplan 1983; 1996).

Joyce (1999: 102–4) and Jeffrey (1984: 138–41) both make similar claims. A particularly detailed argument along these lines for comparative belief can be found in Hawthorne (2009). Indeed, this idea has a long and distinguished history that goes back at least as far as B.O. Koopman (1940). I.J Good (1962), Terrence Fine (1973) and Patrick Suppes (1974) all discussed ideas along these lines. Seidenfeld, Schervish, and Kadane (1995) give a representation theorem for preference that don’t satisfy completeness. (See Evren and Ok 2011; Pedersen 2014; and Chu and Halpern 2008, 2004; for very general representation theorems).

2.3 Weight of evidence, balance of evidence

Evidence influences belief. Joyce (2005) suggests that there is an important difference between the weight of evidence and the balance of evidence. He argues that this is a distinction that precise probabilists struggle to deal with and that the distinction is worth representing. This idea has been hinted at by a great many thinkers including J.M. Keynes, Rudolf Carnap, C.S. Pierce and Karl Popper (see references in Joyce 2005; Gärdenfors and Sahlin 1982). Here’s Keynes’ articulation of the intuition:
As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according as the new knowledge strengthens the unfavourable or the favourable evidence; but something seems to have increased in either case,—we have a more substantial basis upon which to rest our conclusion. I express this by saying than an accession of new evidence increases the weight of an argument. (Keynes 1921: 78, Keynes’ emphasis)

Consider tossing a coin known to be fair. Let’s say you have seen the outcome of a hundred tosses and roughly half have come up heads. Your degree of belief that the coin will land heads should be around a half. This is a case where there is weight of evidence behind the belief.

Now consider another case: a coin of unknown bias is to be tossed. That is, you have not seen any data on previous tosses. In the absence of any relevant information about the bias, symmetry concerns might suggest you take the chance of heads to be around a half. This opinion is different from the above one. There is no weight of evidence, but there is nothing to suggest that your attitudes to $H$ and $T$ should be different. So, on balance, you should have the same belief in both.

However, these two different cases get represented as having the same probabilistic belief, namely $p(H) = p(T) = 0.5$. In the fair coin case, this probability assignment comes from having evidence that suggests that the chance of heads is a half, and the prescription to have your credences match chances (ceteris paribus). In the unknown bias case, by contrast, one arrives at the same assignment in a different way: nothing in your evidence supports one proposition over the other so some “principle of indifference” reasoning suggests that they should be assigned the same credence (see Hájek 2011, for discussion of the principle of indifference).

If we take seriously the “ambiguity aversion” discussed earlier, when offered the choice between betting on the fair coin’s landing heads as opposed to the unknown-bias coin’s landing heads, it doesn’t seem unreasonable to prefer the former. Recall the preference for unambiguous gambles in the Ellsberg game in section 2.1. But if both coins have the same subjective probabilities attached, what rationalises this preference for betting on the fair coin? Joyce argues that there is a difference between these beliefs that is worth representing. IP does represent the difference. The first case is represented by $P(H) = \{0.5\}$, while the second is captured by $P(H) = [0, 1]$.

Scott Sturgeon puts this point nicely when he says:

> Evidence and attitude aptly based on it must match in character. When evidence is essentially sharp, it warrants sharp or exact attitude; when evidence is essentially fuzzy—as it is most of the time—it warrants at best a fuzzy attitude. In a phrase: evidential precision begets attitudinal precision; and evidential imprecision begets attitudinal imprecision. (Sturgeon 2008: 159 Sturgeon’s emphasis)

Wheeler (2014) criticises Sturgeon on this “character matching” thesis. However, an argument for IP based on the nature of evidence only requires that the character of the evidence sometimes allows (or mandates?) imprecise belief and not that the characters must always match. In opposition, Schoenfield (2012) argues that evidence always supports precise credence, but that for reasons of limited computational capacity, real agents needn’t be required to have precise credences. However, her argument only really supports the claim that sometimes indeterminacy is due to complexity of the evidence and computational complexity. She doesn’t have an argument against the claims Levi, Kaplan, Joyce and
others make that there are evidential situations that warrant imprecise attitudes.

Strictly speaking, what we have here is only half the story. There is a difference between the representations of belief as regards weight and balance. But that still leaves open the question of exactly what is representing the weight of evidence? What aspect of the belief reflects this difference? One might be tempted to view \( \overline{P(H)} - \overline{P(H)} \) as a measure of the weight of evidence for \( H \). Walley (1991) tentatively suggests as much. However, this would get wrong cases of conflicting evidence. (Imagine two equally reliable witnesses: one tells you the coin is biased towards heads, the other says the bias is towards tails.) The question of whether and how IP does better than precise probabilism has not yet received an adequate answer. Researchers in IP have, however, made progress on distinguishing cases where your beliefs happen to have certain symmetry properties from cases where your beliefs capture evidence about symmetries in the objects of belief. This is a distinction that the standard precise model of belief fails to capture (de Cooman and Miranda 2007).

The precise probabilist can respond to the weight/balance distinction argument by pointing to the property of resiliency (Skyrms 2011) or stability (Leitgeb 2014). The idea is that probabilities determined by the weight of evidence change less in response to new evidence than do probabilities determined by balance of evidence alone. That is, if you’ve seen a hundred tosses of the coin, seeing it land heads doesn’t affect your belief much, while if you’ve not seen any tosses of the coin, seeing it land heads has a bigger effect on your beliefs. Thus, the distinction is represented in the precise probabilistic framework in the conditional probabilities. The distinction, though, is one that cannot rationalise the preference for betting on the fair coin. One could develop a resiliency-weighted expected value and claim that this is what you should maximise, but this would be as much of a departure from orthodox probabilism as IP is. If someone were to develop such a theory, then its merits could be weighed against the merits of IP type models.

Another potential precise response would be to suggest that there is weight of evidence for \( H \) if many propositions that are evidence for \( H \) are fully believed, or if there is a chance proposition (about \( H \)) that is near to fully believed. This is in contrast to cases of mere balance where few propositions that are evidence for \( H \) are fully believed, or where probability is spread out over a number of chance hypotheses. The same comments made above about resiliency apply here: such distinctions can be made, but this doesn’t get us to a theory that can rationalise ambiguity aversion.

The phenomenon of dilation (section 3.1) suggests that the kind of argument put forward in this section needs more care and further elaboration.

2.4 Suspending judgement

You are sometimes in a position where none of your evidence seems to speak for or against the truth of some proposition. Arguably, a reasonable attitude to take towards such a proposition is suspension of judgement. When there is little or no information on which to base our conclusions, we cannot expect reasoning (no matter how clever or thorough) to reveal a most probable hypothesis or a uniquely reasonable course of action. There are limits to the power of reason. (Walley 1991: 2)

Consider a coin of unknown bias. The Bayesian agent must have a precise belief about the coin’s landing heads on the next toss. Given the complete lack of information about the coin, it seems like it would be better just to suspend judgement. That is, it would be better not to have any particular
precise credence. It would be better to avoid betting on the coin. But there just isn’t room in the Bayesian framework to do this. The probability function must output some number, and that number will sanction a particular set of bets as desirable.

Consider $p(X)$ as representing the degree to which the evidence supports $X$. Now consider $I(X) = 1 - (p(X) + p(\neg X))$. This measures the degree to which the evidence is silent on $X$. Huber (2009) points out that precise probabilism can then be understood as making the claim that no evidence is ever silent on any proposition. That is, $I(X) = 0$ for all $X$. One can never suspend judgement. This is a nice way of seeing the strangeness of the precise probabilist’s attitude to evidence. Huber is making this point about Dempster-Shafer belief functions (see historical appendix, section 7), but it carries over to IP in general.

The committed precise probabilist would respond that setting $p(X) = 0.5$ is suspending judgement. This is the maximally noncommittal credence in the case of a coin flip. More generally, suspending judgement should be understood in terms of maximising entropy (Jaynes 2003; Williamson 2010: 49–72). The imprecise probabilist could argue that this only seems to be the right way to be noncommittal if you are wedded to the precise probabilist representation of belief. That is, the MaxEnt approach makes sense if you are already committed to representation of belief by a single precise probability, but loses its appeal if credal sets are available. Suspending judgement is something you do when the evidence doesn’t determine your credence. But for the precise probabilist, there is no way to signal the difference between suspension of judgement and strong evidence of probability half. This is just the weight/balance argument again.

To make things more stark, consider the following delightfully odd example from Adam Elga:

A stranger approaches you on the street and starts pulling out objects from a bag. The first three objects he pulls out are a regular-sized tube of toothpaste, a live jellyfish, and a travel-sized tube of toothpaste. To what degree should you believe that the next object he pulls out will be another tube of toothpaste? (2010: 1)

In this case, unlike in the coin case, it really isn’t clear what intuition says about what would be the “correct” precise probabilist suspension of judgement. What Maximum Entropy methods recommend will depend on seemingly arbitrary choices about the formal language used to model the situation. Williamson is well aware of this language relativity problem. He argues that choice of a language encodes some of our evidence.

Another response to this argument would be to take William James’ response to W.K. Clifford (Clifford 1901; James 1897). James argued that as long as your beliefs are consistent with the evidence, then you are free to believe what you like. So there is no need to ever suspend judgement. Thus, the precise probabilist’s inability to do so is no real flaw. This attitude, which is sometimes called epistemic voluntarism, is close to the sort of subjectivism espoused by Bruno de Finetti, Frank Ramsey and others.

There does seem to be a case for an alternative method of suspending judgement in order to allow you to avoid making any bets when your evidence is very incomplete, ambiguous or imprecise. If your credences serve as your standard for the acceptability of bets, they should allow for both sides of a bet to fail to be acceptable. A precise probabilist cannot do this since if a bet has (precise) expected value $e$ then taking the other side of that bet (being the bookie) has expected value $-e$. If acceptability is understood as nonnegative expectation, then at least one side of any bet is acceptable to a precise agent. This seems unsatisfactory. Surely genuine
suspension of judgement involves being unwilling to risk money on the truth of a proposition at any odds.

Inspired by the famous “Bertrand paradox”, Chandler (2014) offers a neat argument that the precise probabilist cannot jointly satisfy two desiderata relating to suspension of judgement about a variable. First desideratum: if you suspend judgement about the value of a bounded real variable $X$, then it seems that different intervals of possible values for $X$ of the same size should be treated the same by your epistemic state. Second desideratum: if $Y$ essentially describes the same quantity as $X$, then suspension of judgement about $X$ should entail suspension of judgement about $Y$. Let’s imagine now that you have precise probabilities and that you suspend judgement about $X$. By the first desideratum, you have a uniform distribution over values of $X$. Now consider $Y = 1/X$. $Y$ essentially describes the same quantity that $X$ did. But a uniform distribution over $X$ entails a non-uniform distribution over $Y$. So you do not suspend judgement over $Y$. A real-world case of variables so related is “ice residence time in clouds” and “ice fall rate in clouds”. These are inversely related, but describe essentially the same element of a climate system (Stainforth et al. 2007: 2154).

So a precise probabilist cannot satisfy these reasonable desiderata of suspension of judgement. An imprecise probabilist can: for example, the set of all probability functions over $X$ satisfies both desiderata. There may be more informative priors that also represent suspension of judgement, but it suffices for now to point out that IP seems better able to represent suspension of judgement than precise probabilism. Section 5.5 of Walley (1991), discusses IP’s prospects as a method for dealing with suspension of judgement.

2.5 Unknown correlations

Haenni et al. (2011) motivate imprecise probabilities by showing how they can arise from precise probability judgements. That is, if you have a precise probability for $X$ and a precise probability for $Y$, then you can put bounds on $p(X \cap Y)$ and $p(X \cup Y)$, even if you don’t know how $X$ and $Y$ are related. These bounds give you intervals of possible probability values for the compound events.

For example, you know that $p(X \cap Y)$ is bounded above by $p(X)$ and by $p(Y)$ and thus by $\min\{p(X), p(Y)\}$. If $p(X) > 0.5$ and $p(Y) > 0.5$ then $X$ and $Y$ must overlap. So $p(X \cap Y)$ is bounded below by $p(X) + p(Y) - 1$. But, by definition, $p(X \cap Y)$ is also bounded below by 0. So we have the following result: if you know $p(X)$ and you know $p(Y)$, then, you know

$$\max\{0, p(X) + p(Y) - 1\} \leq p(X \cap Y) \leq \min\{p(X), p(Y)\}.$$  

Likewise, bounds can be put on $p(X \cup Y)$. $p(X \cup Y)$ can’t be bigger than when $X$ and $Y$ are disjoint, so it is bounded above by $p(X) + p(Y)$. It is also bounded above by 1, and thus by the minimum of those expressions. It is also bounded below by $p(X)$ and by $p(Y)$ and thus by their maximum. Putting this together,

$$\max\{p(X), p(Y)\} \leq p(X \cup Y) \leq \min\{p(X) + p(Y), 1\}.$$  

These constraints are effectively what you get from de Finetti’s Fundamental Theorem of Prevision (de Finetti 1990 [1974]: 112; Schervish, Seidenfeld, and Kadane 2008).

So if your evidence constrains your belief in $X$ and in $Y$, but is silent on their interaction, then you will only be able to pin down these compound events to certain intervals. Any choice of a particular probability function will go beyond the evidence in assuming some particular evidential relationship between $X$ and $Y$. That is, $p(X)$ and $p(X | Y)$ will differ in a way that has no grounding in your evidence.
2.6 Nonprobabilistic chances

What if the objective chances were not probabilities? If we endorse some kind of connection between known objective chances and belief—for example, a principle of direct inference or Lewis’ Principal Principle (Lewis 1986)—then we might have an additional reason to endorse imprecise probabilism. It seems to be a truth universally acknowledged that chances ought to be probabilities, but it is a “truth” for which very little argument has been offered. For example, Schaffer (2007) makes obeying the probability axioms one of the things required in order to play the “chance role”, but offers no argument that this should be the case. Joyce says “some have held objective chances are not probabilities. This seems unlikely, but explaining why would take us too far afield” (2009: 279, fn.17). Various other discussions of chance—for example in statistical mechanics (Loewer 2001; Frigg 2008) or “Humean chance” (Lewis 1986, 1994)—take for granted that chances should be precise and probabilistic (Dardashti et al. 2014 is an exception). Obviously things are confused by the use of the concept of chance as a way of interpreting probability theory. There is, however, a perfectly good pre-theoretic notion of chance: this is what probability theory was originally invented to reason about, after all. This pre-theoretic chance still seems like the sort of thing that we should apportion our belief to, in some sense. And there is very little argument that chances must always be probabilities. If the chances were nonprobabilistic in a particular way, one might argue that your credences ought to be nonprobabilistic in the same way. What form a chance-coordination norm should take if chances and credences were to have non-probabilistic formal structures is currently an open problem.

I want to give a couple of examples of this idea. First consider some physical process that doesn’t have a limiting frequency but has a frequency that varies, always staying within some interval. This would be a process that is chancy, but fairly predictable. It might be that the best description of such a system is to just put bounds on its relative frequency. Such processes have been studied using IP models (Kumar and Fine 1985; Grize and Fine 1987; Fine 1988), and have been discussed as a potential source of imprecision in credence (Hájek and Smithson 2012). A certain kind of non-standard understanding of a quantum-mechanical event leads naturally to upper probability models (Suppes and Zanotti 1991; Hartmann and Suppes 2010). John Norton has discussed the limits of probability theory as a logic of induction, using an example which, he claims, admits no reasonable probabilistic attitude (Norton 2007, 2008a,b). One might hope that IP offers an inductive logic along the lines Norton sketches. Norton himself has expressed scepticism on this line (Norton 2007), although Benétreau-Dupin (2015) has defended IP as a candidate system for Norton’s project. Finally, particular views on vagueness might well prompt a rethinking of the formal structure of chance (Bradley 2016).

2.7 Group belief

Suppose we wanted our epistemology to apply not just to individuals, but to “group agents” like committees, governments, companies, and so on. Such agents may be made up of members who disagree. Levi (1986, 1999) has argued that representation of such conflict is better handled with sets of probabilities than with precise probabilities. There is a rich literature on combining or aggregating the (probabilistic) opinions of members of groups (Genest and Zidek 1986) but the outcome of such aggregation does not adequately represent the disagreement among the group. Some forms of aggregation also fail to respect plausible constraints on group belief. For example, if every member of the group agrees that \( X \) and \( Y \) are probabilistically independent, then it seems plausible to require that the group belief respects this unanimity. It is, however, well known that linear pooling—a simple and popular form of aggregation—does not respect this desideratum. Consider two probability functions \( p,q \) such that \( p(X) = p(Y) = 1/3 \) and \( p(X \mid Y) = p(X) \) while \( q(X) = q(Y) = 2/3 \) and
Levi also argued that individual agents can be in conflict in the same way as groups, and thus that individuals’ credal states are also better represented by sets of probabilities. (Levi also argued for the convexity of credal states, which brings him into conflict with the above argument about independence (see historical appendix section 3).) One doesn’t need to buy the claim that groups and individuals must be modelled in the same way to take something away from this idea. One merely needs to accept the idea that an individual can be conflicted in such a way that a reasonable representation of her belief state—or belief and value state—is in terms of sets of functions. Bradley (2009) calls members of such sets “avatars”. This suggests that we interpret an individual’s credal set as a credal committee made up of her avatars. This interpretation of the representor is due to Joyce (2011), though Joyce attributes it Adam Elga. This committee represents all the possible prior probabilities you could have that are consistent with the evidence. Each credal committee member is a fully opinionated Jamesian voluntarist. The committee as a whole, collectively, is a Cliffordian objectivist.

3. Philosophical questions for IP

This section collects some problems for IP noted in the literature.

3.1 Dilation

$q(X | Y) = q(X)$. Consider aggregating these two probabilities by taking an unweighted average of them: $r = p/2 + q/2$. Now, calculation shows that $r(X ∩ Y) = 5/18$ while $r(X)r(Y) = 1/4$, thus demonstrating that $r$ does not consider $X$ and $Y$ to be independent. So such an aggregation method does not satisfy the above desideratum (Kyburg and Pittarelli 1992; Cozman 2012). For more on judgement aggregation in groups, see List and Pettit (2011), especially chapter 2.

Elkin and Wheeler (2016) argue that resolving disagreement among precise probabilist peers should involve an imprecise probability. Stewart and Quintana (2018) argue that imprecise aggregation methods have some nice properties that no precise aggregation method do.

If committee members have credences and utilities that differ among the group, then no precise probability-utility pair distinct from the probabilities and utilities of the agents can satisfy the Pareto condition (Seidenfeld, Kadane, and Schervish 1989). The Pareto condition requires that the group preference respect agreement of preference among the group. That is, if all members of the group prefer $A$ to $B$ (that is, if each group member finds that $A$ has higher expected utility than $B$) then the aggregate preference (as determined by the aggregate probability-utility pair) should satisfy that preference. Since this “consensus preservation” is a reasonable requirement on aggregation, this result shows that precise models of group agents are problematic. Walley discusses an example of a set of probabilities where each probability represents the beliefs of a member of a group, then $P$ is an incomplete description of the beliefs of each agent, in the sense that if all members of $P$ agree on something, then that thing is something each agent believes. Sets of probabilities allow us to represent an agent who is conflicted in their judgements (Levi 1986, 1999).
Consider two logically unrelated propositions $H$ and $X$. Now consider the four “state descriptions” of this simple model as set out in Figure 1. So $a = H \cap X$ and so on. Now define $Y = a \cup d$. Alternatively, consider three propositions related in the following way: $Y$ is defined as “$H$ if and only if $X$”.

\[ \begin{array}{ccc} 
\text{H} & \neg \text{H} & \\
\text{a} & b & X \\
c & d & \neg X \\
\end{array} \]

\text{FIGURE 1: A diagram of the relationships after Seidenfeld (1994); Y is the shaded area}

Further imagine that $p(H \mid X) = p(H) = 1/2$. No other relationships between the propositions hold except those required by logic and probability theory. It is straightforward to verify that the above constraints require that $p(Y) = 1/2$. The probability for $X$, however, is unconstrained.

Let’s imagine you were given the above information, and took your representor to be the full set of probability functions that satisfied these constraints. Roger White suggested an intuitive gloss on how you might receive information about propositions so related and so constrained (White 2010). White’s puzzle goes like this. I have a proposition $X$, about which you know nothing at all. I have written whichever is true out of $X$ and $\neg X$ on the Heads side of a fair coin. I have painted over the coin so you can’t see which side is heads. I then flip the coin and it lands with the $X$ uppermost. $H$ is the proposition that the coin lands heads up. $Y$ is the proposition that the coin lands with the “$X$” side up.

Imagine if you had a precise prior that made you certain of $X$ (this is compatible with the above constraints since $X$ was unconstrained). Seeing $X$ land uppermost now should be evidence that the coin has landed heads. The game set-up makes it such that these apparently irrelevant instances of evidence can carry information. Likewise, being very confident of $X$ makes $Y$ very good evidence for $H$. If instead you were sure $X$ was false, $Y$ would be solid gold evidence of $H$’s falsity. So it seems that $p(H \mid Y)$ is proportional to prior belief in $X$ (indeed, this can be proven rather easily). Given the way the events are related, observing whether $X$ or $\neg X$ landed uppermost is a noisy channel to learn about whether or not $H$ landed uppermost.

So let’s go back to the original imprecise case and consider what it means to have an imprecise belief in $X$. Among other things, it means considering possible that $X$ could be very likely. It is consistent with your belief state that $X$ is such that if you knew what proposition $X$ was, you would consider it very likely. In this case, $Y$ would be good evidence for $H$. Note that in this case learning that the coin landed $\neg X$ uppermost—call this $Y'$—would be just as good evidence against $H$. Likewise, $X$ might be a proposition that you would have very low credence in, and thus $Y$ would be evidence against $H$.

Since you are in a state of ignorance with respect to $X$, your representor contains probabilities that take $Y$ to be good evidence that $H$ and probabilities that take $Y$ to be good evidence that $\neg H$. So, despite the fact that $P(H) = \{1/2\}$ we have $P(H \mid Y) = [0, 1]$. This phenomenon—posteriors being wider than their priors—is known as dilation. The phenomenon has been thoroughly investigated in the mathematical literature (Walley 1991; Seidenfeld and Wasserman 1993; Herron,
Seidenfeld, and Wasserman 1994; Pedersen and Wheeler 2014). Levi and Seidenfeld reported an example of dilation to Good following Good (1967). Good mentioned this correspondence in his follow up paper (Good 1974). Recent interest in dilation in the philosophical community has been generated by White’s paper (White 2010).

White considers dilation to be a problem since learning $Y$ doesn’t *seem* to be relevant to $H$. That is, since you are ignorant about $X$, learning whether or not the coin landed $X$ up doesn’t seem to tell you anything about whether the coin landed heads up. It seems strange to argue that your belief in $H$ should *dilate* from $\frac{1}{2}$ to $[0,1]$ upon learning $Y$. It feels as if this should just be irrelevant to $H$. However, $Y$ is only really irrelevant to $H$ when $p(X) = \frac{1}{2}$. Any other precise belief you might have in $X$ is such that $Y$ now affects your posterior belief in $H$. Figure 2 shows the situation for one particular belief about how likely $X$ is; for one particular $p \in P$. The horizontal line can shift up or down, depending on what the committee member we focus on believes about $X$. $p(H \mid Y)$ is a half only if the prior in $X$ is also a half. However, the imprecise probabilist takes into account *all the ways* $Y$ might affect belief in $H$.

Consider a group of agents who each had precise credences in the above coin case and differed in their priors on $X$. They would all start out with prior of a half in $H$. After learning $Y$, these agents would differ in their posterior opinions about $H$ based on their differing dispositions to update. The group belief would dilate. However, no agent in the group has acted in any way unreasonably. If we take Levi’s suggestion that individuals can be conflicted just like groups can, then it seems that individual agents can have their beliefs dilate just like groups can.

There are two apparent problems with dilation. First, the belief-moving effect of apparently irrelevant evidence; and second, the fact that learning some evidence can cause your belief-intervals to widen. The above comments speak to the first of these. Pedersen and Wheeler (2014) also are focused on mitigating this worry. We turn now to the second worry.

Even if we accept dilation as a fact of life for the imprecise probabilist, it is still *weird*. Even if all of the above argument is accepted, it still seems strange to say that your belief in $H$ is dilated, *whatever you learn*. That is, whether you learn $Y$ or $Y'$, your posterior belief in $H$ looks the same: $[0,1]$. Or perhaps, what it shows to be weird is that your initial credence was precise.

Hart and Titelbaum (2015) suggest that dilation is strange because conditionalising on a biconditional (which is, after all, what you are doing in the above example) is unintuitive even in the precise case. Whether all cases of dilation can be explained away in this manner remains to be seen. Gong and Meng (2017) likewise see dilation as a problem of mis-specified statistical inference, rather than a problem for IP per se.

Beyond this seeming strangeness, White suggests a specific way that being subject to dilation is an indicator of a defective epistemology. White suggests that dilation examples show that imprecise probabilities violate

![Figure 2: A member of the credal committee (after Joyce (2011))](image-url)
the Reflection Principle (van Fraassen 1984). The argument goes as follows:

given that you know now that whether you learn $Y$ or you learn $Y'$
your credence in $H$ will be $[0,1]$ (and you will certainly learn one
or the other), your current credence in $H$ should also be $[0,1]$.

The general idea is that you should set your credences to what you expect
your credences to be in the future. More specifically, your credence in $X$
should be the expectation of your future possible credences in $X$ over the
things you might learn. Given that, for all the things you might learn in
this example your credence in $H$ would be the same, you should have that
as your prior credence also. Your prior should be such that $P(H) = [0,1]$
So having a precise prior credence in $H$ to start with is irrational. That’s
how the argument against dilation from reflection goes. Your prior $P$ is not
fully precise though. Consider $P(H \cap Y)$. That is, the prior belief in the
conjunction is imprecise. So the alleged problem with dilation and
reflection is not as simple as “your precise belief becomes imprecise”.
The problem is “your precise belief in $H$ becomes imprecise”; or rather, your
precise belief in $H$ as represented by $P(H)$ becomes imprecise.

The issue with reflection is more basic. What exactly does reflection
require of imprecise probabilists in this case? Now, it is obviously the case
that each credal committee member’s prior credence is its expectation over
the possible future evidence (this is a theorem of probability theory). But
somehow, it is felt, the credal state as a whole isn’t sensitive to reflection
in the way the principle requires. Each $p \in P$ satisfies the principle, but
the awkward symmetries of the problem conspire to make $P$ as a whole
violate the principle. This looks to be the case if we focus on $P(H)$ as an
adequate representation of that part of the belief state. But as noted earlier,
this is not an adequate way of understanding the credal state. Note that
while learning $Y$ and learning $Y'$ both prompt revision to a state where the
posterior belief in $H$ is represented as an interval by $[0,1]$, the credal
states as sets of probabilities are not the same. Call the state after learning
$Y$, $P'$ and the state after learning $Y'$, $P''$. So $P' = \{p(\cdot) \mid Y \}, p \in P\}$ and
$P'' = \{p(\cdot) \mid Y'\}, p \in P\}$. While it is true that $P'(H) = P''(H), P' \neq P''$ as
sets of probabilities, since if $p \in P'$ then $p(Y) = 1$ whereas if $p \in P''$ then
$p(Y) = 0$. So one lesson we should learn from dilation is that imprecise
belief is represented by sets of functions rather than by a set-valued
function (see also, Joyce 2011; Topey 2012; Bradley and Steele 2014b).

So dilation can perhaps be tamed or rationalised, and the issue with
reflection can be mitigated. But there is still a puzzle that dilation raises: in
the precise context we have a nice result – due to Good (1967) – that says
roughly that learning new information has positive expected value.
Information has positive value. This result is, to some extent, undermined
by dilation. Bradley and Steele (2016) suggest that there is some sense in
which Good’s result can be partially salvaged in the IP setting.

It seems that examples of dilation undermine the earlier claim that
imprecise probabilities allow you to represent the difference between the
weight and balance of evidence (see section 2.3): learning $Y$ appears to
give rise to a belief which one would consider as representing less
evidence since it is more spread out. This is so because the prior credence
in the dilation case is precise, not through weight of evidence, but through
the symmetry discussed earlier. We cannot take narrowness of the interval
$[P(X), \overline{P}(X)]$ as a characterisation of weight of evidence since the interval
can be narrow for reasons other than because lots of evidence has been
accumulated. So my earlier remarks on weight/balance should not be read
as the claim that imprecise probabilities can always represent the
weight/balance distinction. What is true is that there are cases where
imprecise probabilities can represent the distinction in a way that impacts
on decision making. This issue is far from settled and more work needs to
be done on this topic.
3.2 Belief inertia

Imagine there are two live hypotheses $H_1$ and $H_2$. You have no idea how likely they are, but they are mutually exclusive and exhaustive. Then you acquire some evidence $E$. Some simple probability theory shows that for every $p \in P$ we have the following relationship (using $p_i = p(E \mid H_i)$ for $i = 1, 2$).

$$p(H_1 \mid E) = \frac{p(E \mid H_1)p(H_1)}{p_1p(H_1) + p_2p(H_2)}$$

$$= \frac{p_1}{p_1 + (p_1 - p_2)p(H_1)}$$

If your prior in $H_1$ is vacuous—if $P(H_1) = [0, 1]$—then the above equation shows that your posterior is vacuous as well. That is, if $p(H_1) = 0$ then $p(H_1 \mid E) = 0$ and likewise for $p(H_1) = 1 = p(H_1 \mid E)$, and since the right hand side of the above equation is a continuous function of $p(H_1)$, for every $r \in [0, 1]$ there is some $p(H_1)$ such that $p(H_1 \mid E) = r$. So $P(H_1 \mid E) = [0, 1]$.

It seems like the imprecise probabilist cannot learn from vacuous priors. This problem of belief inertia goes back at least as far as Levi (1980), chapter 13. Walley also discusses the issue, but appears unmoved by it: he says that vacuous posterior probabilities are just a consequence of adopting a vacuous prior:

The vacuous previsions really are rather trivial models. That seems appropriate for models of "complete ignorance" which is a rather trivial state of uncertainty. On the other hand, one cannot expect such models to be very useful in practical problems, notwithstanding their theoretical importance. If the vacuous previsions are used to model prior beliefs about a statistical parameter for instance, they give rise to vacuous posterior previsions… However, prior previsions that are close to vacuous and make nearly minimal claims about prior beliefs can lead to reasonable posterior previsions. (Walley 1991: 93)

Joyce (2011) and Rinard (2013) have both discussed this problem. Rinard’s solution to it is to argue that this shows that the vacuous prior is never a legitimate state of belief. Or rather, that we only ever need to model your beliefs using non-vacuous priors, even if these are incomplete descriptions of your belief state. This is similar to Walley’s “non-exhaustive” representation of belief. Vallinder (forthcoming) suggests that the problem of belief inertia is quite a general one. Castro and Hart use the looming danger of belief inertia to argue against what I have called an "objectivist" interpretation of IP.

An alternative solution to this problem, (inspired by Wilson 2001; and Cattaneo 2008; 2014), would modify the update rule in such a way that those extreme priors that give extremely small likelihoods to the evidence are excised from the representer. More work would need to be done to make this precise and show how exactly the response would go.

3.3 Decision making

One important use that models of belief can be put to is as part of a theory of rational decision. IP is no different. Decision making with imprecise probabilities has some problems, however.

The problem for IP decision making, in short, is that your credal committee can disagree on what the best course of action is, and when they do, it is unclear how you should act (recall the definitions in section 1.1). Imagine betting on a coin of unknown bias. Consider the indicator gambles on heads and tails. Both bets have imprecise expectation $[0, 1]$. How are you supposed to compare these expectations? The bets are incomparable. (If the coin case appears to have too much exploitable
symmetry, consider unit bets on Elga pulling toothpaste or jellyfish from
his bag.) This incomparability, argues Williamson, leads to decision
making paralysis, and this highlights a flaw in the epistemology (2010: 70). This argument seems to be missing the point, however, if one of our
motivations for IP is precisely to be able to represent such incompatibility
of prospects (see section 2.2)! The incommensurability of options entailed
by IP is not a bug, it’s a feature. Decision making with imprecise
probabilities is discussed by Seidenfeld (2004), Troffaes (2007),
Seidenfeld, Schervish, and Kadane (2010), Bradley (2015), Williams

A more serious worry confronts IP when you have to make sequences of
decisions. There is a rich literature in economics on sequences of decisions
for agents who fail to be orthodox expected utility maximisers (Seidenfeld
1988; 1994; Machina 1989; Al-Najjar and Weinstein 2009, and the
references therein). This topic was brought to the attention of philosophers
again after the publication of Elga’s (2010) paper Subjective Probabilities
Should Be Sharp which highlights the problem with a simple decision
example, although a very similar example appears in Hammond (1988) in
relation to Seidenfeld’s discussion of Levi’s decision rule “E-
admissibility” (Seidenfeld 1988).

A version of the problem is as follows. You are about to be offered two
bets on a coin of unknown bias, A and B, one after the other. The bets pay
out as follows:

- A loses 10 if the coin lands heads and wins 15 otherwise
- B wins 15 if the coin lands heads and loses 10 otherwise

If we assume you have beliefs represented by \( P(H) = [0, 1] \), these bets
have expectations of \([-10, 15]\). Refusing each bet has expectation of 0. So
accepting and refusing A are incomparable with respect to your beliefs.

Likewise for B. The problem is that refusing both bets seems to be
irrational, since accepting both bets gets you a guaranteed payoff of 5.
Elga argues that no decision rule for imprecise probabilities can rule out
refusing both bets. He then argues that this shows that imprecise
probabilities are bad epistemology. Neither argument works. Chandler
(2014) and Sahlin and Weirich (2014) both point out that a certain kind of
imprecise decision rule does make refusing both bets impermissible and
Elga has acknowledged this in an erratum to his paper. Bradley and Steele
(2014a) argue that decision rules that make refusing both bets merely
permissible are legitimate ways to make imprecise decisions. They also
point out that the rule that Chandler, and Sahlin and Weirich advocate has
counterintuitive consequences in other decision problems.

Moss (2015) relates Elga-style IP decision problems to moral dilemmas
and uses the analogy to explain the conflicting intuitions in Elga’s
problem. Sud (2014) and Rinard (2015) both also offer alternative
decision theories for imprecise probabilities. Bradley (2019) argues that all
three struggle to accommodate a version of the Ellsberg decisions
discussed above.

Even if Elga’s argument worked and there were no good imprecise
decision rules, that wouldn’t show that IP was a faulty model of belief. We
want to be able to represent the suspension of judgement on various
things, including on the relative goodness of a number of options. Such
incommensurability inevitably brings with it some problems for sequential
decisions (see, for example, Broome 2000), but this is not an argument
against the epistemology. As Bradley and Steele note, Elga’s argument—if
it were valid—could mutatis mutandis be used as an argument that there
are no incommensurable goods and this seems too strong.

3.4 Interpreting IP
Imprecise probabilities aren’t a radically new theory. They are merely a slight modification of existing models of belief for situations of ambiguity. Often your credences will be precise enough, and your available actions will be such that you act more or less as if you were a strict Bayesian. One might analogize imprecise probabilities as the “Theory of Relativity” to the strict Bayesian “Newtonian Mechanics”: all but indistinguishable in all but the most extreme situations. This analogy goes deeper: in both cases, the theories are “empirically indistinguishable” in normal circumstances, but they both differ radically in some conceptual respects. Namely, the role of absolute space in Newtonian mechanics/GR; how to model ignorance in the strict/imprecise probabilist case. Howson (2012) makes a similar analogy between modelling belief and models in science. Both involve some requirement to be somewhat faithful to the target system, but in each case faithfulness must be weighed up against various theoretical virtues like simplicity, computational tractability and so on. Likewise Hosni (2014) argues that what model of belief is appropriate is somewhat dependent on context. There is of course an important disanalogy in that models of belief are supposed to be normative as well as descriptive, whereas models in science typically only have to play a descriptive role. Walley (1991) discusses a similar view but is generally sceptical of such an interpretation.

3.4.1 What is a belief?

One standard interpretation of the probability calculus is that probabilities represent “degrees of belief” or “credences”. This is more or less the concept that under consideration so far. But what is a degree of belief? There are a number of ways of cashing out what it is that a representation of degree of belief is actually representing.

One of the most straightforward understandings of degree of belief is that credences are interpreted in terms of an agent’s limiting willingness to bet.

This is an idea which goes back to Ramsey (1926) and de Finetti (1964, 1990 [1974]). The idea is that your credence in \( X \) is \( \alpha \) just in case \( \alpha \) is the value at which you are indifferent between the gambles:

- Win \( 1 - \alpha \) if \( X \), lose \( \alpha \) otherwise
- Lose \( 1 - \alpha \) if \( X \), win \( \alpha \) otherwise

This is the “betting interpretation”. This is the interpretation behind Dutch book arguments: this interpretation of belief makes the link between betting quotients and belief strong enough to sanction the Dutch book theorem’s claim that beliefs must be probabilistic. Williamson in fact takes issue with IP because IP cannot be given this betting interpretation (2010: 68–72). He argues that Smith’s and Walley’s contributions notwithstanding (see formal appendix), the single-value betting interpretation makes sense as a standard for credence in a way that the one-sided betting interpretation doesn’t. The idea is that you may refuse all bets unless they are at extremely favourable odds by your lights. Such behaviour doesn’t speak to your credences. However, if you were to offer a single value then this tells us something about your epistemic state. There is something to this idea, but it must be traded off against the worry that forcing agents to have such single numbers systematically misrepresents their epistemic states. As Kaplan puts it

The mere fact that you nominate 0.8 under the compulsion to choose some determinate value for \( \rho(X) \) hardly means that you have a reason to choose 0.8. The orthodox Bayesian is, in short, guilty of advocating false precision. (Kaplan 1983: 569, Kaplan’s emphasis)

A related interpretation of credence is to understand credence as being just a representation of an agent’s dispositions to act. This interpretation sees credence as that function such that your elicited preferences and observed
actions can be represented as those of an expected utility maximiser with respect to that probability function (Briggs 2014: section 2.2). Your credences just are that function that represents you as a rational agent. For precise probabilism, “rational agent” means “expected utility maximiser”. For imprecise probabilism, rational agent must mean something slightly different. A slightly more sophisticated version of this sort of idea is to understand credence to be exactly that component of the preference structure that the probability function represents in the representation theorem. Recall the discussion of incompleteness (section 2.2). IP represents you as the agent conflicted between all the $p \in P$ such that unless the $p$ agree that $X$ is better than $Y$ or vice versa, you find them incomparable. What a representation theorem actually proves is a matter of some dispute (see Zynda 2000; Hájek 2008; Meacham and Weisberg 2011).

One might take the view that credence is modelling some kind of mental or psychological quantity in the head. Strength of belief is a real psychological quantity and it is this that credence should measure. Unlike the above views, this interpretation of credence isn’t easy to operationalise. It also seems like this understanding of strength of belief distances credence from its role in understanding decision making. The above 

behaviourist 

views take belief’s role in decision making to be central to or even definitional of what belief is. This psychological interpretation seems to divorce belief from decision. Whether there are such stable neurological structures is also a matter of some controversy (Fumagalli 2013; Smithson and Pushkarskaya 2015).

A compromise between the behaviourist views and the psychological views is to say that belief is characterised in part by its role in decision making. This leaves room for belief to play an important role in other things, like assertion or reasoning and inference. So the answer to the question “What is degree of belief?” is: “Degree of belief is whatever psychological factors play the role imputed to belief in decision making contexts, assertion behaviour, reasoning and inference”. There is room in this characterisation to understand credence as measuring some sort of psychological quantity that causally relates to action, assertion and so on. This is a sort of functionalist reading of what belief is. Eriksson and Hájek (2007) argue that “degree of belief” should just be taken as a primitive concept in epistemology. The above attempts to characterise degree of belief then fill in the picture of the role degree of belief plays.

3.4.2 What is a belief in $X$?

So now we have a better idea of what it is that a model of belief should do. But which part of our model of belief is representing which part of the belief state? The first thing to say is that $P(X)$ is not an adequate representation of the belief in $X$. That is, one of the values of the credal set approach is that it can capture certain kinds of non-logical relationships between propositions that are lost when focusing on, say, the associated set of probability values. For example, consider tossing a coin of unknown bias. $P(H) = P(T) = [0, 1]$, but this fails to represent the important fact that $p(H) = 1 - p(T)$ for all $p \in P$. Or that getting a heads on the first toss is at least as likely as heads on two consecutive tosses. These facts that aren’t captured by the sets-of-values view can play an important role in reasoning and decision.

$P(X)$ might be a good enough representation of belief for some purposes. For example in the Ellsberg game these sets of probability values (and their associated sets of expectations) are enough to rationalise the non-probabilistic preferences. How good the representation needs to be depends on what it will be used for. Representing the sun as a point mass is a good enough representation for basic orbital calculations, but obviously inadequate if you are studying coronal mass ejections, solar
flares or other phenomena that depend on details of the internal dynamics of the sun.

3.5 Regress

Imprecise probabilities is a theory born of our limitations as reasoning agents, and of limitations in our evidence base. If only we had better evidence, a single probability function would do. But since our evidence is weak, we must use a set. In a way, the same is true of precise probabilism. If only we knew the truth, we could represent belief with a truth-valuation function, or just a set of sentences that are fully believed. But since there are truths we don’t know, we must use a probability to represent our intermediate confidence. And indeed, the same problem arises for the imprecise probabilist. Is it reasonable to assume that we know what set of probabilities best represents the evidence? Perhaps we should have a set of sets of probabilities... Similar problems arise for theories of vagueness (Sorensen 2012). We objected to precise values for degrees of belief, so why be content with sets-valued beliefs with precise boundaries? This is the problem of “higher-order vagueness” recast as a problem for imprecise probabilism. Why is sets of probabilities the right level to stop the regress at? Why not sets of sets? Why not second-order probabilities? Why not single probability functions? Williamson (2014) makes this point, and argues that a single precise probability is the correct level at which to get off the “uncertainty escalator”. Williamson advocates the betting interpretation of belief, and his argument here presupposes that interpretation. But the point is still worth addressing: for a particular interpretation of what belief is, what sort of level of uncertainty is appropriate. For the functionalist interpretation suggested above, this is something of a pragmatic choice. The further we allow this regress to continue, the harder it is to deal with these belief-representing objects. So let’s not go further than we need.

We have seen arguments above that IP does have some advantage over precise probabilism, in the capacity to represent suspending judgement, the difference between weight and balance of evidence and so on. So we must go at least this far up the uncertainty escalator. But for the sake of practicality we need not go any further, even though there are hierarchical Bayes models that would give us a well-defined theory of higher-order models of belief. This is, ultimately, a pragmatic argument. Actual human belief states are probably immensely complicated neurological patterns with all the attendant complexity, interactivity, reflexivity and vagueness.

We are modelling belief, so it is about choosing a model at the right level of complexity. If you are working out the trajectory of a cannonball on earth, you can safely ignore the gravitational influence of the moon on the cannonball. Likewise, there will be contexts where simple models of belief are appropriate: perhaps your belief state is just a set of sentences of a language, or perhaps just a single probability function. If, however, you are modelling the tides, then the gravitational influence of the moon needs to be involved: the model needs to be more complex. This suggests that an adequate model of belief under severe uncertainty may need to move beyond the single probability paradigm. But a pragmatic argument says that we should only move as far as we need to. So while you need to model the moon to get the tides right, you can get away without having Venus in your model. This relates to the contextual nature of appropriateness for models of belief mentioned earlier. If one were attempting to provide a complete formal characterisation of the ontology of belief, then these regress worries would be significantly harder to avoid.

Let’s imagine that we had a second order probability \( \mu \) defined over the set of (first order) probabilities \( P \). We could then reduce uncertainty to a single function by \( p^*(X) = \sum_P \mu(p)p(X) \) (if \( P \) is finite, in the interests of keeping things simple I discuss only this case). Now if \( p^*(X) \) is what is used in decision making, then there is no real sense in which we have a genuine IP model, and it cannot rationalise the Ellsberg choice, nor can it
give rise to incomparability. If there is some alternative use that \( \mu \) is put to, a use that allows incomparability and that rationalises Ellsberg choices, then it might be a genuine rival to credal sets, but it represents just as much of a departure from the orthodox theory as IP does.

Gärdenfors and Sahlin’s *Unreliable Probabilities* model enriches a basic IP approach with a “reliability index” (see the historical appendix). Lyon (2017) enriches the standard IP picture in a different way: he adds a privileged “best guess” probability. This modification allows for better aggregation of elicited IP estimates. How best to interpret such a model is still an open question. Other enriched IP models are no doubt available.

### 3.6 What makes a good imprecise belief?

There are, as we have seen, certain structural properties that are necessary conditions on rational belief. What exactly these are depends on your views. However, there are further ways of assessing belief. Strongly believing true things and strongly believing the negations of false things seem like good-making-features of beliefs. For the case of precise credences, we can make this precise. There is a large literature on “scoring rules”: methods for measuring how good a probability is relative to the actual state of the world (Brier 1950; Savage 1971; Joyce 2009; Pettigrew 2011). These are numerical methods of measuring how good a probability is given the true state of the world.

For the case of imprecise probabilities, however, the situation looks bleak. No real valued scoring rule for imprecise probabilities can have the desirable property of being *strictly proper* (Seidenfeld, Schervish, and Kadane 2012). Schoenfield (2017) presents a simple version of the result. Since strict propriety is a desirable property of a scoring rule (Bröcker and Smith 2007; Joyce 2009; Pettigrew 2011), this failing is serious. So further work is needed to develop a well-grounded theory of how to assess imprecise probabilities. Mayo-Wilson and Wheeler (2016) provide a neater version of the proof, and offer a property weaker than strict propriety that an imprecise probability scoring rule can satisfy. Carr (2015) and Konek (forthcoming) both present positive suggestions for moving forward with imprecise scoring rules. Levinstein (forthcoming) suggests that the problem really only arises for determinately imprecise credences, but not for indeterminate credence.

### 4. Summary

Imprecise probabilities offer a model of rational belief that does away with some of the idealisation required by the orthodox precise probability approach. Many motivations for such a move have been put forward, and many views on IP have been discussed. There are still several open philosophical questions relating to IP, and this is likely to be a rich field of research for years to come.

### Bibliography


—. 2019, “A counterexample to three imprecise decision theories”, Theoria, 85:1 18–30
—. 2016, “Can free evidence be bad? Value of information for the imprecise probabilist”, Philosophy of Science, 83:1 1–28
Clifford, William Kingdom, 1901, “The Ethics of Belief”, in Lectures and Essays, Leslie Stephen and Frederick Pollock (eds), 2:161–205, 3rd
Meeting of the Philosophy of Science Association, 250–259.


Konek, Jason, forthcoming “Epistemic conservativity and imprecise credence”, Philosophy and Phenomenological Research


——, 2008a, “Ignorance and Indifference”, *Philosophy of Science*, 75: 45–68.


Smithson, Michael, and Helen Pushkarskaya, 2015, “Ignorance and the


Weichselberger, Kurt, 2000, “The theory of interval-probability as a unifying concept for uncertainty”, International Journal of...


Acknowledgments

Many thanks to Teddy Seidenfeld, Greg Wheeler, Paul Pedersen, Aidan Lyon, Catrin Campbell-Moore, Stephan Hartmann, the ANU Philosophy of Probability Reading Group, and an anonymous referee for helpful comments on drafts of this article.
Some formal machinery was introduced in section 1.1. Refer back to that section for definitions. Recall that the objects of belief are an algebra of subsets of some set of states $\Omega$.

- 1. Properties of strength of belief
- 2. The betting justification of IP
- 3. From lower probabilities to credal sets

1. Properties of strength of belief

We are interested in strength of belief and thus in some sort of graded notion of belief. One straightforward and popular way to represent strength of belief is with a function that maps the objects of belief to real numbers: bigger numbers represent more strongly believed sentences. Let’s consider a function, $b$, from subsets of $\Omega$ to real numbers. In this section I outline a few properties we might want these numerical representations of belief to satisfy (see also Huber 2014).

By convention, let $b(\Omega) = 1$ and $b(\emptyset) = 0$. It seems reasonable to suppose that your beliefs in other sentences are bounded above and below by these expressions: it would be odd if there were something you believed more strongly than the necessary event, or less strongly than the impossible event. Call this property $-b(\emptyset) \leq b(X) \leq b(\Omega)$ for all $X$—boundedness.

If $X$ is always true whenever $Y$ is, then it seems sensible to require that you believe $X$ at least as much as you do $Y$. Call this property—if $Y \subseteq X$ then $b(Y) \leq b(X)$—monotonicity.

Consider the following property: If $X$ and $Y$ are incompatible—if $X \cap Y = \emptyset$—then $b(X \cup Y) \geq b(X) + b(Y)$. This property is called superadditivity. If the "$\geq$" is replaced by a "$\leq$" we have the property of subadditivity.

A slight strengthening of superadditivity gives us 2-monotone: $b(X \cup Y) \geq b(X) + b(Y) - b(X \cap Y)$. It is straightforward to show that this entails superadditivity, and only a little more work to show that the converse is false.

A further strengthening of this idea is $n$-monotone. This is slightly more complicated than the previous ones.

$$b\left(\bigcup_{i=1}^{n} X_i\right) \geq \sum_{i=1}^{n} \sum_{\mathcal{I} \subseteq \{1, \ldots, n\}, |\mathcal{I}| = i} (-1)^{i+1} \left(\bigcap_{j \in \mathcal{I}} X_j\right)$$

Let’s break this down. On the left hand side, we are constraining belief in unions of $n$ events. How are we constraining them? Look at the inner summation on the right hand side. This is the sum of all the $i$-event intersections. If $i$ is even, we take each of these away, if $i$ is odd, we add them. The outer sum ranges over $i$. So we add the one member intersections, take away the two member intersections, add the three element intersections, and so on up to $n$. If $n = 2$ then we have exactly the formula we had above for 2-monotone. If $b$ is $n$-monotone, then it is $(n - 1)$-monotone. If $b$ is $n$-monotone for all $n$, then $b$ is called infinite monotone or totally monotone.

One final property is additivity which results from replacing the inequality in the description of superadditivity with an equality: if $X$ and $Y$ are incompatible, then $b(X \cup Y) = b(X) + b(Y)$. A bounded, monotonic function on an algebra of events is called a capacity. A bounded, infinite monotone function on an algebra of events is called a belief function. A bounded, additive function on an algebra of events is a probability function. Belief functions and capacities represent one way of going beyond orthodox probability theory.
One might want to move beyond orthodox probability theory by relaxing the constraint that degree of belief be represented by a real valued function. One might instead take belief to be represented by a function that maps events to intervals of the real line, or to sets of real numbers (Weichselberger 2000). These approaches can, to some extent, be subsumed under the more general way of thinking about imprecise belief using sets of probabilities that we encountered in section 1.1, where your credal state is represented by a set of probability functions \( P \). Recall that your lower envelope of \( X \) is: \( \underline{P}(X) = \inf_{P}(X) \); and your upper envelope is \( \overline{P}(X) = \sup_{P}(X) \). \( \underline{P} \) is a superadditive capacity, \( \overline{P} \) is a subadditive capacity.

This brief survey doesn’t come close to mapping out the whole range of possible representations of uncertainty (but see Halpern 2003; Haenni et al. 2011; Augustin et al. 2014; Huber 2014) There is, however, a philosophical tradition that puts sets of probabilities and belief functions foremost, and this article is primarily about that tradition. Credal sets are also a very general theory that subsumes many of the other representations. So in what follows the discussion will be in terms of sets of probability functions.

One might wonder whether taking representors to be sets of conditional probabilities as the basic item might not be a better way to do things. Hájek (2003) argues that conditional probabilities should be taken to be the basic entity, and unconditional probabilities should be defined out of them. In what follows, this point won’t really make any difference, so I will continue to talk mostly of unconditional probabilities. Bear in mind that these are “really” conditional probabilities conditioned on the tautology.

2. The betting justification of IP

Bruno de Finetti justified that your degrees of belief should have the structure of a probability function in the following way. Belief is understood as a summary of your attitudes to gambles. That is, belief is what determines your dispositions to choose among bets, and thus determines your betting behaviour. He showed that if you satisfy some reasonable constraints on how your attitudes to gambles are structured, then your degrees of belief are, in a certain sense, probabilistic (de Finetti 1964; 1990 (1974)).

The insight behind de Finetti’s approach was that your attitudes to how likely various events are could be read off your attitudes to gambles. Recall that gambles are bounded real valued functions (Troffaes and de Cooman 2014) extend the theory to unbounded gambles). Of particular interest are the indicator functions that output 1 if the proposition in question is true and 0 otherwise.

De Finetti essentially appealed to the claim that if you are unwilling to buy a bet at a particular price, you must be happy to sell the bet for that price. This was one of his criteria of “coherence”. C.A.B. Smith questioned this assumption (Smith 1961) and Peter Walley built an extremely powerful and rich theory on the basis of these insights (Walley 1991). For a self-contained summary of this "lower previsions" approach, see Miranda (2008). For a modern book-length treatment of the theory see Troffaes and de Cooman (2014).

There’s quite a lot of set up required here, but the punchline is worth it. If we assume that you are neither risk seeking nor risk averse, and that the gambles are denominated in some currency such that your utility is linear in that currency, then your attitude about the contingencies can be read off your attitude about gambles in a straightforward way. Before we explain this, we add some structure to the set of gambles. This presentation follows Quaeghebeur (2014). Call \( \mathcal{L} \) the set of bounded real valued
functions on $\Omega$. If $f, g \in \mathcal{L}$ are gambles, then so is their “pointwise sum” $(f + g)(\omega) = f(\omega) + g(\omega)$ for all $\omega$. For a real number $\lambda$, if $f$ is a gamble then so is $(\lambda f)(\omega) = \lambda f(\omega)$ for all $\omega$. Let real number $\lambda$ stand also for the constant gamble that pays out $\lambda$: $\lambda(\omega) = \lambda$ for all $\omega$.

Consider the set of desirable gambles $D$. These are gambles that you would be willing to accept if they were offered to you. Which gambles you would willingly accept reflects your opinions in the following way: the more you would be willing to pay for a gamble $f$, the more likely you consider the contingencies where $f$ gets you a good prize. Consider the indicator function $I_X$ for an event $X$. This is the function that outputs 1 if $\omega$ is in $X$ and 0 otherwise. The more you would be willing to pay for the gamble $I_X$, the more likely you take $X$ to be. In what follows I will use $X$ interchangeably for the event and its indicator function. Being willing to pay $\alpha$ for the gamble $f$ is the same as finding the gamble $f - \alpha$ acceptable. Consider the following function: $\bar{E}(f) = \sup\{\alpha \in \mathbb{R}, f - \alpha \in D\}$. Call this your “lower prevision” for $f$. That is, your lower prevision for the gamble $f$ is the largest amount you would pay to receive the gamble $f$. We can also define $\overline{E}(f) = \inf\{\alpha \in \mathbb{R}, \alpha - f \in D\}$: the smallest amount you’d be willing to sell the gamble $f$ for. These two functions are conjugate in the sense that $\overline{E}(f) = -\overline{E}(-f)$. We will be particularly interested in lower previsions of indicator functions: these directly reflect how likely you think the event in question is. $X - \alpha$ is effectively a unit bet on $X$ that costs you $\alpha$. De Finetti focussed on the case where if $f - \alpha \not\in D$ then $\alpha - f \in D$ for all $f$ and all $\alpha$. Then $\overline{E}(f) = \overline{E}(f)$ and we can describe this number—call it $E(f)$—as your “fair price” for the gamble $f$. It’s the price that makes you indifferent between buying and selling gamble $f$. Smith (1961) allowed your fair selling price and your fair buying price to differ.

Here’s how we now proceed. We describe some reasonable “coherence constraints” on sets of desirable gambles, and then describe what structure such constraints put on $\overline{E}, D$ should be closed under addition. That is, if $f, g \in D$ then $f + g \in D$. Or, more succinctly, $D + D \subseteq D$. $D$ should also be closed under multiplication by a positive constant $\lambda > 0$: if $f \in D$ then $\lambda f \in D$, or more succinctly $\lambda D = D$. Call $L^+$ the set of gambles that are always positive, that is $f \in L^+$ iff $f(\omega) > 0$ for all $\omega$. Likewise $L^-$ is the set of everywhere negative gambles. Our third condition is that always positive gambles should be in $D$: $L^+ \subseteq D$. And fourth, no always negative gambles should be in $D$: $L^- \cap D = \emptyset$.

Now, $D$ is a coherent set of gambles—i.e., it satisfies the above four properties—if and only if $\overline{E}$ has the following structure:

- $E(f) \geq \inf_{\omega} \{f(\omega), \omega \in \Omega\}$
- $\overline{E}(f + g) \geq \overline{E}(f) + \overline{E}(g)$
- if $\lambda > 0$ then $\overline{E}(\lambda f) = \lambda \overline{E}(f)$

Call a lower prevision coherent if and only if it satisfies these properties. Call a set of gambles $D$ maximal when $f \in D$ iff $-f \not\in D$ for all $f \not\in 0$. $\overline{E} = \overline{E}$ if $D$ is coherent and maximal. The above is, strictly speaking, not quite true. There are some subtleties about topological properties of sets of desirable gambles—whether bets on the boundary of $D$ are in $D$—that preclude the link being quite this straightforward, but these need not detain us (Quaeghebeur 2014). When $\overline{E} = \overline{E}$, we call this function, $E$, a linear prevision because such previsions satisfy the following properties:

- $E(f) \geq \inf_{\omega} \{f(\omega), \omega \in \Omega\}$
- $E(f + g) = E(f) + E(g)$
- and therefore $E(\lambda f) = \lambda E(f)$ for all $\lambda \in \mathbb{R}$

Now, here is the punchline: if we restrict attention to linear previsions of indicator functions, we get exactly the probability functions! (De Finetti didn’t use the “sets of desirable gambles” approach to coherence, but rather something almost equivalent to it, see Quaeghebeur (2014) and...
Miranda and de Cooman (2014) (for details). The same move in the case of lower previsions—restricting attention to indicator functions—yields lower probabilities. As well as Smith, P.M. Williams was another author to consider weakening de Finetti’s requirement in 1975, although the paper was published only in 2007 (Williams 2007; Vicig, Zaffalon, and Cozman 2007). Building on Smith’s and Williams’ insight, Walley (1991) builds a rich and sophisticated theory of statistical inference on lower previsions and lower probabilities. See Augustin et al. (2014) for a modern introduction to IP, and Vicig and Seidenfeld (2012) for a more careful discussion of the history of IP. Because of the conjugacy of the upper and lower probabilities—because \( \overline{P}(X) = 1 - \widetilde{P}(\neg X) \)—the lower probability tells us all we need to know about your belief state. In fact, coherent \( P \) and \( \overline{P} \) are related as described in Figure F1 (Walley 1991: 84 ff).

![Figure F1: Relationships between upper and lower probabilities](image)

An arrow between two expressions means that the higher one is greater than or equal to the lower one. The solid lines highlight the nontrivial inequalities. If \( A \) and \( B \) are incompatible, it is clear that this tangle of inequalities simplifies to:

\[
\overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B) \geq \overline{P}(A) + \overline{P}(B) \geq P(A \cup B) \geq P(A) + P(B)
\]

3. From lower probabilities to credal sets

To see how the above foundation relates to the credal sets approach, we need a final couple of formal tools. Say that prevision \( \mathcal{E} \) dominates \( \mathcal{E}' \) if, for all \( f \in \mathcal{E}, \mathcal{E}(f) \geq \mathcal{E}'(f) \). For a lower prevision \( \mathcal{E} \), define the envelope of \( \mathcal{E} \) as the set of linear previsions that dominate \( \mathcal{E} \). Call this \( M(\mathcal{E}) \). For a set of linear previsions \( \mathcal{M} \), define their lower envelope as \( \inf\{E(f), E \in \mathcal{M}\} \). The envelope theorem states that \( \mathcal{E} \) is coherent iff \( \mathcal{E} \) is the lower envelope of \( M(\mathcal{E}) \). Restricted to the case of lower previsions of indicator functions—lower probabilities—this states that your betting behaviour is coherent if and only if your fair buying prices are the lower envelope of a set of probabilities.

This puts the imprecise probabilities approach on almost the same footing as the precise probability approach in that we have an exact analogue of de Finetti’s version of the so-called Dutch book theorem: the theorem has less demanding—more reasonable—premises, and the conclusion is that rational agents’ beliefs should have the structure of a lower probability and thus through the envelope theorem, the structure of a credal set. One disanalogy between de Finetti’s result and the IP one-sided result is that your precise fair prices determine your probability function whereas in the one-sided betting case, the coherence requirements only constrain your representor. That is, a complete set of coherent fair prices pick out a unique probability, while several distinct sets of probabilities can give rise to the same set of one-sided previsions.

Historical appendix: Theories of imprecise belief

In this section we review some authors who held views that were or can be interpreted as IP friendly. This list is not exhaustive: these authors are those who have been influential or who had particularly interesting theories. These sections offer mere sketches of the often rich and interesting theories that have been put forward.
J.M. Keynes

In his *Treatise on Probability* Keynes argued that “probabilities” needn’t always be amenable to numerical comparisons (Keynes 1921). He said:

[N]o exercise of the practical judgment is possible, by which a numerical value can actually be given to the probability of every argument. So far from our being able to measure them, it is not even clear that we are always able to place them in an order of magnitude. Nor has any theoretical rule for their evaluation ever been suggested. The doubt, in view of these facts, whether any two probabilities are in every case even theoretically capable of comparison in terms of numbers, has not, however, received serious consideration. There seems to me to be exceedingly strong reasons for entertaining the doubt. (Keynes 1921: 29)

I maintain … that there are some pairs of probabilities between the members of which no comparison of magnitude is possible; that we can say, nevertheless, of some pairs of relations of probability that the one is greater and the other less, although it is not possible to measure the difference between them; and that in a very special type of case … a meaning can be given to a numerical comparison of magnitude. (Keynes 1921: 36, Keynes’ emphasis)

Imprecise Probabilities

Stanford Encyclopedia of Philosophy

Keynes’ *Theory of Probability* contains the diagram reproduced in Figure H1, and it’s clear from this that he thought there could be degrees of belief that were not numerically comparable. Keynes interprets the $O$ and $I$ as the contradiction and tautology respectively and $A$ is a proposition with a numerically measurable probability. The lines connect those propositions (denoted by letters) that can be compared. So $V$ and $W$ can be compared and $W$ is more likely that $V$ (since it is closer to $I$). Those propositions without lines between them (for example $X$ and $Y$) are incomparable. Keynes’ own discussion of the diagram is on page 42 of Keynes (1921).

Weatherson (2002) interprets Keynes as favouring some sort of IP view since sets of functions (or intervals of values) naturally give rise to the sorts of incomparabilities that Keynes takes to be features of belief. Keynes took (conditional) probability to be a sort of logical relationship that held between propositions (Hájek 2011: 3.2), rather than as strength of belief. So whether Keynes would have approved of IP models is unclear. See Kyburg (2003) for a discussion of Keynes’ view by someone sympathetic to IP, and see Brady and Arthmar (2012) for further discussion of Kaynes’ view.

2. I.J. Good

I.J. Good, mathematician, statistician, student of G.H. Hardy and Bletchley Park code-breaker, was an early advocate of something like IP.
In principle these [probability] axioms should be used in conjunction with inequality judgements and therefore only lead to inequality discernments... Most subjective probabilities are regarded as belonging only to some interval of values. (Good 1983 [1971]: 15)

Good is usually associated with the “black box model of belief”. The idea here is that at the core of your epistemic state is a “black box” which you cannot look inside. The black box outputs “discernments”: qualitative probability judgements like “X is more likely than Y”. The idea is that inside the black box there is a numerical and precise probability function that you only have indirect and imperfect access to (Good 1962).

It isn’t wholly clear from Good’s writings on this topic whether the precise probability in the black box is supposed to be a genuine part of an agent’s epistemic state or whether talk of the precise black box probability is just a calculational device. Good is often interpreted as being in the former camp (by Levi for example), but the following quote suggests he might have been in the latter camp:

It is often convenient to forget about the inequality for the sake of simplicity and to simply use precise estimates. (Good 1983 [1971]: 16)

The way Good talks in his (1962)—especially around page 77—also suggests that Good’s view isn’t quite the “black box” interpretation that is attributed to him. In any case, Good is certainly interpreted as holding the view that IP is required since belief is only imperfectly available to introspection.

3. Isaac Levi

Isaac Levi has been a prominent defender of a particular brand of IP since the seventies (Levi 1974, 1980, 1985, 1986). Levi’s brand of IP is developed carefully and thoroughly over the course of several books and articles, the most important being The Enterprise of Knowledge (Levi 1980) and Hard Choices (Levi 1986). Levi has several motivations for being dissatisfied with the precise probabilistic orthodoxy. One of Levi’s motivations for IP is captured by the following quote:

[It] is sometimes rational to make no determinate probability judgement and, indeed, to make maximally indeterminate judgements. Here I am supposing ... that refusal to make a determinate probability judgement does not derive from a lack of clarity about one’s credal state. To the contrary, it may derive from a very clear and cool judgement that on the basis of the available evidence, making a numerically determinate judgement would be unwarranted and arbitrary. (Levi 1985: 396)

Another motivation for Levi’s brand of IP is connected to his general picture of the structure of an agent’s mental state at a time. Levi’s view is that your belief state at a time can be captured by two components. The first component is your Corpus of Knowledge, \( K \), which “encodes the set of sentences ... to whose certain truth [you are] committed” (Levi 1974: 395). \( K \) is a deductively closed set of sentences. In any model of partial belief and uncertain inference, there are some things about which you are uncertain. But in every case there are some things that are taken for granted. Imagine a toy example of drawing marbles from an urn. The observed frequencies of colours is used as evidence to infer something about the frequencies of colours in the urn. In this model, you take for granted that you are accurately recognising the colours and are not being deceived by an evil demon or anything like that (or, more prosaically, that the draws are probabilistically independent of each other and each draw has the same probabilities associated with it, i.e. that the draws are
independent and identically distributed, i.i.d.). That’s not to say that we couldn’t model a situation where there was some doubt about the observations: the point is that in the simple case, that sort of funny business is just ruled out. “No funny business” is in $K$, if you like. There are certain aspects of the situation that are taken for granted: that are outside the modelling process. This is the same in science: when we model the load stress of a particular design of bridge, we take for granted the basic physics of rigid bodies and the facts about the materials involved.

The second part of Levi’s view of your belief state is captured by your confirmational commitments which describe how you are disposed to change your belief in response to certain kinds of evidence. \( C \) is a function that takes your corpus $K$ as input and outputs “your beliefs”: some object that informs your decision making behaviour and constrains how your attitudes change in response to evidence. Levi argues in favour of this bipartite epistemology as follows:

One of the advantages of introducing confirmational commitments is that confirmational commitments and states of full belief can vary independently without incoherence... [T]he separability of confirmational commitment and state of full belief is important to the study of conditions under which changes in credal state are justified. If this separability is ignored and attention is focused on changes in credal state directly, the distinction between changes in credal state that are changes in confirmational commitment and changes in full belief is suppressed without argument. (Levi 1999: 515)

Levi sees precise Bayesianism as suppressing this distinction. The only way a Bayesian agent updates is through conditionalisation. Levi has more to say on the subject of the defects of Bayesianism, but we don’t have room to discuss his criticisms.

For Levi, the output of $C(K)$ is a set of probability functions, $P$, with the following properties:

- $P$ is non-empty.
- $P$ is convex, meaning if $p, q \in P$ then for $0 \leq \lambda \leq 1$, $\lambda p + (1 - \lambda)q \in P$.
- If you learn $E$ with certainty and nothing else, then $C(K + E) = Q$ where $Q = \{p(\cdot | E)\}$.

The first and third of these properties shouldn’t strike anyone as particularly surprising or unusual. The second, however, needs further comment. Recall that Levi thought that sets of probabilities were useful as a way of representing conflict. If you are conflicted between $p$ and $q$ (represented by both being in your representor) then any convex combination of $p$ and $q$ will:

have all the earmarks of potential resolutions of the conflict; and, given the assumption that one should not preclude potential resolutions when suspending judgement between rival systems... all weighted averages of the two [probability] functions are thus taken into account. (Levi 1980: 192).

Levi’s reasons for taking linear averages of precise credal states (and only linear averages) to be particularly pertinent as resolutions of conflict aren’t particularly clearly spelled out. Levi does appeal to a theorem from Blackwell and Girshick (1954) and to Harsanyi’s theorem (Harsanyi 1955) as reasons to think that conflicts in values (utility functions) should be resolved through linear averaging (Levi 1980: 175; Levi 1986: 78). The premises of these arguments are not particularly discussed or justified in Levi’s set up. The argument for credal convexity is less clear, but see (Levi 1980: chap. 9). A further puzzle for Levi’s view is that these potential...
resolutions of conflict might not respect probabilistic independencies that $p$ and $q$ agree on (see section 2.7)

4. Henry Kyburg

Henry Kyburg’s view on rational belief bears some similarities to Levi’s (both were students of Ernest Nagel). Both take there to be some corpus of knowledge $K$—Kyburg’s term is Evidential Corpus—which is part of what determines your credal state. However, Kyburg’s Evidential Probabilities are quite different to Levi’s picture of rational belief.

Kyburg’s $K$ is a collection of sentences of a logical language that includes the expressive resources to say things like “the proportion of $Fs$ that are $Gs$ is between $l$ and $u$”. Such evidential corpora are indexed by a significance level at which the claims in the corpus are valid. Levi doesn’t say a great deal about how the confirmational commitment $C$ constrains your credal state except to say that if $X$ is a sentence in $K$ then if $p \in C(K)$ then $p(X) = 1$. Kyburg has a lot more to say about this step. He develops a set of rules for dealing with conflicts of information in the corpus.

Evidential probabilities are a kind of “interval-valued” probability together with a thorough theory for inferring them from statistical data, and for logically reasoning about them. The intervals aren’t necessarily interpretable as sets of probability functions: that is, they can violate some of the properties of credal sets discussed in the formal appendix. The theory is most fully set out in Kyburg and Teng (2001) (see also, Wheeler and Williamson (2011); Haenni et al. (2011)). The theory is discussed more with an eye to the concerns of psychologists and experimental economists in Kyburg (1983).

5. The SIPTA community

Drawing on the work of Bruno de Finetti, C.A.B. Smith and Peter Williams, there is a rich tradition of IP research that culminates in the body of work developed by those associated with SIPTA: the Society for Imprecise Probability, Theory and Applications (http://www.sipta.org/). SIPTA only came into being in 2002, and the ISIPTA conferences have only been running since 1999, but one can trace a common thread of research back much further. Among the important works contributing to this tradition are: Fine (1973), Suppes (1974), Williams (1976), Walley and Fine (1982), Kyburg (1987), Seidenfeld (1988), Walley (1991), Kadane, Schervish, and Seidenfeld (1999), de Cooman and Miranda (2007), Troffaes and de Cooman (2014) and Augustin et al. (2014). One main strand of work in this tradition is outlined in the formal appendix. See Miranda (2008) for a recent survey of some work in this tradition, and Augustin et al. (2014) for a book length introduction. A recent comment on the history of imprecise probabilities can be found in Vicig and Seidenfeld (2012). Work in what I am calling the “SIPTA tradition” isn’t just focussed on IP as a model of rational belief, but on IP as a useful tool in statistics and other disciplines as well: IP as a model of some non-deterministic process, IP as a tool for classification of data...

The popularity of the term “Imprecise Probability” for the class of models we are interested in is due, in large part, to the influence of Peter Walley’s 1991 book Statistical Reasoning with Imprecise Probabilities. This book was, until very recently, the most complete description of the theory of imprecise probabilities. Walley brought together and extended the above mentioned results and produced a rich and deep theory of statistics on the basis of IP. Despite being mainly devoted to the exposition of the formal machinery of IP, Walley’s book contains a lot of interesting material on the philosophical foundations of IP. In particular, sections 1.3–1.7, the sections on the interpretation of IP (2.10, 2.11), and chapter 5 all contain interesting philosophical discussion that in many ways anticipates recent philosophical debates on IP. It wouldn’t be uncharitable to describe a lot of
recent philosophical work on IP as footnotes to Walley (although referencing Walley in a footnote appears to be as close as some authors get to actually engaging with him). Engagement by philosophers with this community has sadly been rather limited, excepting Levi, Kyburg and their students. This is unfortunate since many philosophically rich topics emerge in the course of this sort of research, for example, the many distinct independence concepts for IP (Kyburg and Pittarelli 1992; Cozman and Walley 2005; Cozman 2012), the rational requirements on group belief (Seidenfeld, Kadane, and Schervish 1989) or the distinction between symmetries in the model and models constrained to satisfy certain symmetries (de Cooman and Miranda 2007).

6. Richard Jeffrey

Richard Jeffrey is sometimes taken to be someone whose views are consonant with the IP tradition. In reality, Jeffrey’s view is a little more complicated. In The Logic of Decision (Jeffrey 1983), Jeffrey develops a representation theorem based on mathematical work by Ethan Bolker that uses premises that are interestingly weaker than those of Savage’s classic theorem (Savage 1972 [1954]). Agents still have complete and transitive preferences, but they have those preferences over a space where the “belief” and “value” parts aren’t straightforwardly separable. In Savage’s theorem, in contrast, the “states” and “outcomes” are distinct spaces. The representation that arises in Jeffrey’s framework is not unique, in the following sense. If \((p, v)\) is a probability-utility representation of the preference relation, then there exists a \(\lambda\) such that \((p', v')\) also represents the preference, where \(p'\) and \(v'\) are defined as:

\[
\begin{align*}
  p'(X) &= p(X)[1 - \lambda v(X)] \\
  v'(X) &= v(X)[(1 + \lambda)/(1 + \lambda v(X))]
\end{align*}
\]

Indeed, there will be infinitely many such representations (Jeffrey 1983: chap. 8; Joyce 1999: 133–5). Jeffrey argued that an epistemology built on such a representation theorem gives

one clear version of Bayesianism in which belief states... are naturally identified with infinite sets of probability functions, so that degrees of belief in particular propositions will normally be determined only up to an appropriate quantization, i.e., they will be interval-valued (so to speak). (Jeffrey 1984)

Jeffrey claims that his theory subsumes Levi’s theory. Levi (1985) responds that his theory and Jeffrey’s are importantly distinct, and Jeffrey (1987) recants.

This aspect of Jeffrey’s work—the Jeffrey-Bolker representation theorem—cannot be taken as a basis for imprecise probabilities in the sense we are considering. Jeffrey notes this point:

[T]he indeterminacy of \([p]\) and \([v]\) that is implied by Bolker’s uniqueness theorem takes place within a context of total determinacy about preference and indifference. Thus it has nothing to do with the decision-theoretical questions that Levi addresses. (Jeffrey 1987: 590)

The indeterminacy of the belief in this setting is due to the inseparability of the beliefs and the values (as captured by the above mentioned alternative representations). However, Jeffrey also takes a line that is more IP-friendly. He says:

I do not take belief states to be determined by full preference rankings of rich Boolean algebras of propositions, for our actual preference rankings are fragmentary... [E]ven if my theory were like Savage’s in that full rankings of whole algebras always
Jeffrey’s main concern in his 1984 paper is with scientific reasoning, and with a solution to the Problem of Old Evidence originally introduced by Glymour (1980). However, with respect to decision theory, he seems much less certain of the role of IP. In places, Jeffrey seems to agree with Levi: “Jeffrey here seems to be suggesting that IP is at least permissible.” However, he ends up thinking that applying the principle of indifference (in some form) is a legitimate way to “sharpen up” your beliefs and values (and thus your preferences). After likening his position to Kaplan’s—see section 2.2—Jeffrey says:

I differ from Kaplan, who would see my adoption of the uniform distribution as unjustifiable precision, whereas I think I would adopt it as a precise characterization of my uncertainty. Having more hopes for locally definite judgmental probabilities than Levi does, I am less dedicated to judgmental sets [of probabilities] as characterizations of uncertainty… I attribute less psychological immediacy than Levi does to sets of judgmental probability functions; and in decision theory, where he assigns them a central systematic role, I use them peripherally and ad hoc. (1987: 589)

Dempster–Shafer belief theory builds a theory of rational belief on an infinite monotone capacity \( bel \) and its conjugate \( plaus(X) = 1 - bel(\neg X) \) (see the formal appendix). \( bel(X) \) is interpreted as the degree to which your evidence supports \( X \). The interesting aspect of DS theory is its method for combining different bodies of evidence. So if \( bel_1 \) represents the degree to which a certain body of evidence supports various hypotheses and \( bel_2 \) captures the degree of support of another body of evidence, then DS theory gives you a method for working out the degree to which the combination of the evidence supports hypotheses. This is a distinct process from conditionalisation which also has an analogue in DS theory (see Kyburg 1987 for discussion of the difference). In any case, DS theory can be, in a sense, subsumed within the credal sets approach, since every DS belief function is the lower envelope of some set of probabilities (Theorem 2.4.1 on p.34 of Halpern 2003). Discussing the details would take us too far afield, so I point the interested reader to the following references: (Halpern 2003: 32–40, 92–5; Kyburg and Teng 2001: 104–12; Haenni 2009: 132–47; Huber 2014: section 3.1).

Gärdenfors and Sahlin (1982) introduce a theory that they term *Unreliable Probabilities* (the theory builds on Gärdenfors 1979). It bears some resemblance to Kyburg’s theory—they in fact discuss Kyburg—but it is a
theory built with decision making in mind. The basic idea is that you have a set of probabilities and attached to each probability function is an indicator of its reliability. Depending on the circumstances, you pick some reliability threshold $\rho$ and restrict your attention to the set of probabilities that are at least as reliable as that threshold. They then have a story about how decision making should go with this set. Note that they don’t really need a measure of reliability, all they need is something to order the probabilities in $P$. The threshold then becomes some cut-off probability: anything less reliable than it doesn’t make the cut.

Gärdenfors and Sahlin (1982) don’t really discuss this measure of reliability a great deal. Presumably reliability increases as evidence comes in that supports that probability function. Gärdenfors and Sahlin offer an example to illustrate how reliability is supposed to work. They consider three tennis matches. In match A, you know that the two tennis players are of roughly the same level and that it will be a tight match. In match B, you have never even heard of either player and so cannot judge whether or not they are well matched. In match C you have heard that the players are really unevenly matched: one player is much better than the other, but you do not know which of the players is significantly better. If we graphed reliability of a probability against how likely that probability thinks it is that the player serving first will win, the graphs would be as follows: graph $A$ would be very sharply peaked about 0.5; graph $B$ would be quite spread out; graph $C$ would be a sort of “U” shape with high reliability at both ends, lower in the middle. See Figure H2. Graph $A$ is peaked because you know that the match will be close. You have reliable information that the player serving has about a 50% chance of winning. Graph $B$ is spread out because you have no such information in this case. In case $C$, you know that the probability functions that put the chances at near 50-50 are unreliable: all you know is that the match will be one-sided.

In summary, the unreliable probabilities approach enriches a credal set with a reliability index. See Levi (1982) for a critical discussion of unreliable probabilities. Cattaneo (2008) gives some content to the reliability index of a probability by interpreting it in terms of the likelihood of the evidence given by that probability. Another formal theory that seems inspired by the reliability approach is that of Brian Hill (2013).

Copyright © 2019 by the author
Seamus Bradley