Respecting evidence: belief functions not imprecise probabilities

Nicholas J.J. Smith[1]

Abstract

The received model of degrees of belief represents them as probabilities. Over the last half century, many philosophers have been convinced that this model fails because it cannot make room for the idea that an agent's degrees of belief should respect the available evidence. In its place they have advocated a model that represents degrees of belief using imprecise probabilities (sets of probability functions). This paper presents a model of degrees of belief based on Dempster–Shafer belief functions and then presents arguments for belief functions over imprecise probabilities as a model of evidence-respecting degrees of belief. The arguments cover three kinds of issue: theoretical virtues (simplicity, interpretability and flexibility); motivations; and problem cases (dilation and belief inertia).

Keywords Degree of belief · Evidence · Probability · Imprecise probability · Dempster–Shafer belief function · Dilation · Belief inertia

1 Introduction

The received model of degrees of belief represents them as probabilities. Over the last half century, however, many philosophers have been convinced that this model fails because it cannot allow that degrees of belief *respect the evidence*—and in its place they have advocated a model that represents degrees of belief using imprecise

---

✉ Nicholas J.J. Smith
nicholas.smith@sydney.edu.au

[1] Philosophy, The University of Sydney, Main Quadrangle A14, Sydney, NSW 2006, Australia

probabilities.[1] This paper presents a model of degrees of belief based on Dempster–Shafer belief functions and argues that it does better than the imprecise probability model at accounting for the idea that degrees of belief should respect the evidence.

The paper proceeds as follows. Belief functions have not received a lot of attention in the philosophical literature, so Sect. 2 presents them in detail: not just the formal details but also a new way of picturing belief functions that should make them more accessible to intuition. After that, I present arguments for belief functions over imprecise probabilities as a model of evidence-respecting degrees of belief. The arguments focus on three kinds of issue. First, theoretical virtues. Section 3 argues that belief functions generally provide simpler models than imprecise probabilities. Section 4 argues that, while the imprecise probability framework lacks a workable story about how to extract information about an agent's belief state from the formal model—about how the model relates to the phenomena—there is, in contrast, a simple and direct interpretation of belief function models. Section 5 argues that the belief function framework has the flexibility to model an important view about how degrees of belief can respond to certain kinds of evidence—whereas the imprecise probability model cannot faithfully model this view. Second, motivations. In the subsequent two sections I critique arguments for the imprecise probability framework. I discern two main lines of argument in the literature. Section 6 argues that the first fails. Section 7 argues that the second lends at least as much support to the belief function view. Third, problem cases. Section 8 considers two well-known problems for the imprecise probability view—dilation and belief inertia—and shows that the belief function view avoids these problems.

2 Belief functions

Van Fraassen (1989, p. 161) writes:

> To depict your state of opinion, you can use a model which I call the Muddy Venn Diagram. Just represent the propositions you have an opinion about by areas on the usual Venn diagram, and represent your personal probability by means of some quantity of mud heaped on them…Call the total mud present

---

[1] See Sect. 3 for quotations and references. On terminology: In the philosophical literature, the term 'imprecise probability' is generally used in connection with a particular kind of model of belief states: one which models them as *sets* of (precise) probability functions. (See e.g. the opening sentence of Bradley & Steele (2014): "Imprecise probabilism—the view that your belief or credal state is best represented by a *set of probability functions*—has received a lot of attention recently.") Another term sometimes used in this connection in the philosophical literature is 'mushy credence'. In the literature in computer science, artificial intelligence, expert systems, engineering, statistics and elsewhere, the term 'imprecise probability' is used as an umbrella term for non-probabilistic models of uncertainty, in which uncertainty is modelled by something other than a single, precise probability function. (See e.g. Walley (2006, p. 3353): "*Imprecise probability* is used as a generic term to cover all mathematical models which measure chance or uncertainty without sharp numerical probabilities.") There are many such models, including (to name just some, and to cite only a few key works) upper and lower probabilities (Smith, 1961), Dempster–Shafer belief functions (Dempster, 1967, 1968; Shafer, 1976), sets of probability functions (i.e. imprecise probabilities in the first sense noted above) (Levi, 1974, 1980; Kaplan, 1983, 1996; van Fraassen, 1990; Joyce, 2010), upper and lower previsions (Walley, 1991; Troffaes & de Cooman, 2014), and probability intervals (Kyburg & Teng, 2001). In this paper I use 'imprecise probability' in the first, philosophical sense.
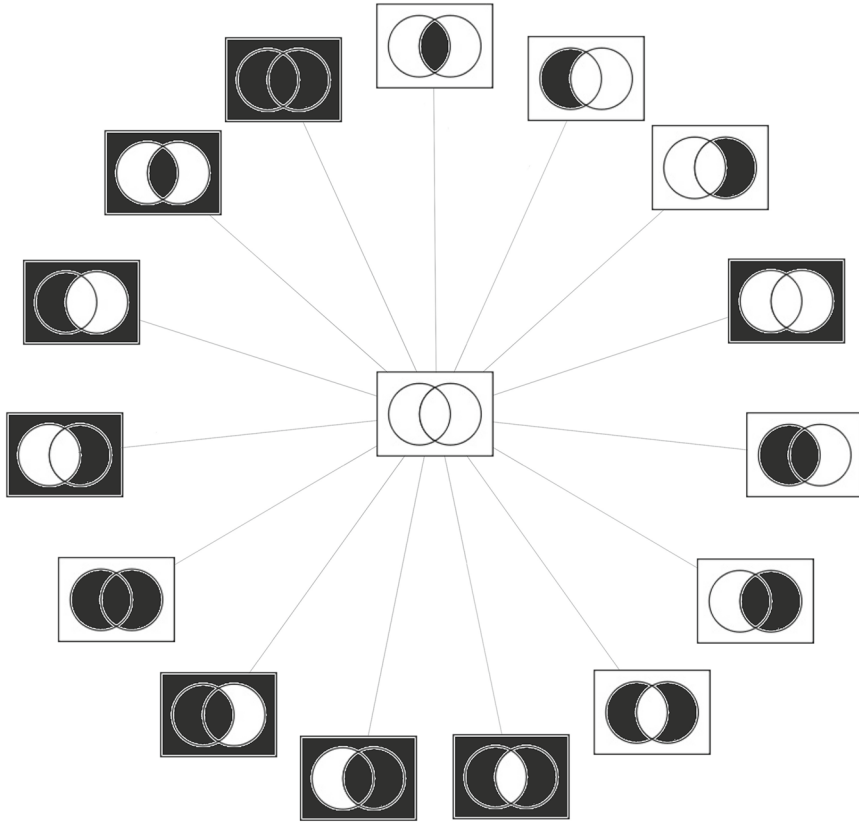
Fig. 1 An exploded Venn diagram

one unit always; the proportion of mud on an area equals the probability of the proposition represented by that area.

We can picture degrees of belief modelled by belief functions in a similar way but with one crucial difference: instead of a Venn diagram we picture what I'll call an *exploded* Venn diagram. Where van Fraassen has us spread our unit of belief over a Venn diagram such as the one at the centre of Fig. 1, my new picture involves our apportioning our unit of belief to its non-empty subsets considered individually—that is, the black shaded regions. (It is because I consider the subsets individually, rather than fitted together into the Venn diagram, that I use the term 'exploded'—by analogy with exploded parts diagrams of the kind found in technical manuals.)

When we move to exploded Venn diagrams, the idea of representing the unit of belief by something spreadable like mud loses much of its point. We may just as well think of it as something rigid like a metal rod of unit mass which can be crosscut into sections of arbitrary length, each of which is then like a weight of the kind used on pan balances. The image then is that we distribute slices of the rod to the exploded black regions. Indeed you can even picture the lines connecting the black regions to
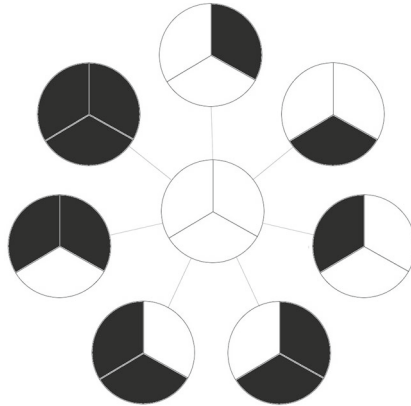
Fig. 2 An exploded non-Venn diagram

the central Venn diagram as scale arms, and the black regions as pans, so that the exploded Venn diagram is like a set of scales with one pan per nonempty subset. You distribute a rod of unit mass—cut into cross-sections of whatever length you please—to these pans. You can put the whole rod (or, if you still prefer, all of the mud) on one region—or a bit here and a bit there.

Venn diagrams are useful when the space of possible outcomes is generated by a finite set of basic (logically atomic) propositions (or propositional variables). Starting with n such propositions generates 2^n state descriptions: one per assignment of truth values to the basic propositions. These constitute the cells of a partition of the space: they correspond to the simple (not further subdividable) regions shown in the Venn diagram.[2] But the core idea also applies when the space of possible outcomes is given in some other way: for example, when it is partitioned into six cells, each corresponding to one of the possible ways a die could fall; or three cells, each corresponding to one of the suspects in a case being considered by a detective. In the latter case, van Fraassen's idea would extend to having us spread mud over the (non Venn) diagram at the centre of Fig. 2, while my new proposal would have us place weights on the black shaded regions.

The reader may object that it would be too mentally taxing to have to consider (say) the 63 separate nonempty subsets of a six-element sample space and apportion weight to each of them—whereas one can easily spread mental mud over a six-section grid. But this is a mistake. Suppose the evidence warrants assigning a certain weight to each specific outcome: say $\frac{1}{6}$ to each section of the six-section grid.[3] Then one assigns $\frac{1}{6}$ to each one-element subset and 0 to everything else: this is just as easy in the exploded case as it is in the unexploded case. In general, if the evidence warrants a particular probabilistic spreading of mud with amount a_i going to outcome A_i, then

---

[2] For example in the Venn diagram at the centre of Fig. 1, if the left and right circles represent the propositions A and B then the four simple regions are the outer 'binocular view surroundings' (A and B both false), the central 'convex lens' (A and B both true), the left 'moon' (A true and B false) and the right 'moon' (A false and B true).

[3] E.g. each section represents one face of a die and the evidence is that the die is fair.

in the exploded case one simply places weight a_i on each singleton subset $\{A_i\}$ and places nothing on any other subset.[4] But suppose the evidence tells us only, say, that the perpetrator is twice as likely to be male than female. Then in the exploded case one can just assign $\frac{2}{3}$ of the weight to the subset containing male suspects and $\frac{1}{3}$ to the subset containing female suspects: simple! Whereas in the probability case you cannot assign $\frac{2}{3}$ to the set of males without also assigning a specific amount to *each* male (and similarly for females). But you may have no evidence at all that tells you how to do that.[5] So it is never harder, and often easier, to place weights on the pans in the exploded diagram than to spread mud over the unexploded diagram.

The story about distributing parts of a unit weight to an exploded diagram gives an intuitive picture of belief functions. Now for the formal details. We begin with a space Ω of mutually exclusive and jointly exhaustive possible outcomes. I shall take Ω to be finite. This is not mathematically essential (see e.g. Shafer (1979)) but it is simpler. Yet simplicity is not my reason for assuming finitude. Rather, it seems that Ω must be finite when it is supposed—as it is in this paper—to distinguish possible outcomes insofar as you could have *evidence* for one but not the other, or would potentially *act* differently if you took one rather than the other to be the case. The argument for this conclusion is essentially the same as Turing's (1936, §1, §9) argument that a person engaged in mechanical computation can make use of only finitely many symbols and can be in one of only finitely many different internal states. While from a metaphysical point of view there are presumably infinitely many possible worlds, from the point of view of an agent gathering and weighing evidence and deciding how to act, the relevant space Ω of possible outcomes will be only finitely differentiable. If, for example, the agent is offered an infinite lottery, or faces a decision situation in which the outcomes vary according to the value of a real-valued variable such as temperature, the relevant sample space will be an (extremely large) finite one that 'chunks' the possibilities at a point where the agent is no longer able to tell which one has occurred (she cannot measure temperature with infinite precision) or to care which one has occurred (she does not have the mental capacity to desire differently every one of an infinite number of possible dollar amounts of prize money).[6] Following Shafer (1976, p. 36) I call Ω the *frame of discernment*.

A *basic mass assignment*[7] is a function $m : 2^\Omega \to [0, 1]$ satisfying two conditions:[8]

$$1.\ m(\emptyset) = 0$$
$$2.\ \sum_{A \subseteq \Omega} m(A) = 1 \tag{1}$$

---

[4] Cf. the discussion of Bayesian belief functions below.

[5] See Sect. 7.

[6] I acknowledge that some readers may find this contentious—and in response, note again that the finitude assumption is not essential.

[7] Terms used in the literature include 'basic probability assignment' (Shafer, 1976, p. 38) 'basic probability number' (Smets, 1981), 'basic belief assignment' (Smets & Kennes, 1994, p. 196) and 'mass function' (Denœux, 2019, p. 93).

[8] 2^Ω denotes the set of all subsets of Ω. It can also be denoted $\wp(\Omega)$.

The exploded diagram image depicts m. $m(A)$ is the amount of mass placed on subset A and the entire unit mass (= condition 2) must be distributed among the nonempty (= condition 1) subsets.[9]

Given a basic mass assignment m the corresponding *belief function* $Bel : 2^\Omega \to [0, 1]$ is defined as follows (for all $A \subseteq \Omega$):

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

Bel is the proposed model of the agent's degrees of belief. The degree of belief in a subset A of the frame of discernment is the sum of the basic masses assigned to all subsets of A (including A itself). In terms of the exploded diagram image, it is the sum of the masses assigned to shapes that do not protrude outside the A shape when the exploded pieces are assembled.

The sets to which m assigns a nonzero value are called the *focal sets* of m. If the focal sets of m are all singletons, the belief function Bel corresponding to m is a probability function—that is:

1. $Bel(\emptyset) = 0$
2. $Bel(\Omega) = 1$ (3)
3. $Bel(A \cup B) = Bel(A) + Bel(B)$ whenever $A, B \subseteq \Omega$ and $A \cap B = \emptyset$

A belief function that satisfies these three conditions is called *Bayesian*.

The difference between the probability and belief function frameworks is thus as follows. In both cases one assigns a number between 0 and 1 to each subset A of the sample space, representing one's degree of belief that the true outcome is in A. In the probability case these assignments are all determined by an assignment of portions of a unit of mass to *singleton* subsets (or equivalently elements) of the sample space. The degree of belief of an arbitrary subset A is then the sum of the masses assigned to A's singleton subsets (or members). The belief function approach can be described by removing the word 'singleton'. We start with an assignment of portions of a unit of mass to *subsets* of the sample space. The final or overall degree of belief in a subset A is then the sum of the masses assigned to A's subsets. Both the probability and belief function frameworks thus involve a distinction between fundamental and derived quantities. The fundamental probabilities are the assignments to singleton sets (or equivalently elements). The derived probabilities are the assignments to arbitrary subsets (derived from the fundamental probabilities by addition). In the belief function framework, the basic masses (the values of the basic mass assignment m) are the fundamental assignments. The degrees of belief (the values of the belief function Bel) are the derived assignments (derived from the basic assignments by Eq. (2)).[10]

The idea is that a basic mass assignment is warranted or induced by a body of evidence. For example, if Ω is a set of suspects {Alice, Bob, Carol, Dave, Edwina,

---

[9] Condition 1 can be dropped: see Smets (1988). I shall not explore this route here.

[10] Equation (2) involves addition (we add m values to get Bel values) but note that, unlike probability functions, belief functions are not (in general) additive (in the sense of condition 3 in (3)). See below for further discussion of additivity and the relationship between probability functions and belief functions.

Frank} then the testimony of a witness that a woman left the building will warrant a basic mass assignment that assigns some positive mass p to {Alice, Carol, Edwina} and the rest (i.e. $1 - p$) to Ω.[11] The testimony of a witness that Alice left the building will warrant an assignment of q to {Alice} and $1 - q$ to Ω. The discovery that the suspects rolled a fair die to determine who would commit the crime leads to an assignment of $\frac{1}{6}$ to each singleton subset of Ω (which induces a Bayesian belief function). A completely empty body of evidence induces the assignment of 1 to Ω; and so on. The belief function generated from the mass function then represents overall degree of belief. The overall degree of belief that the perpetrator was a woman, say, will be not only the mass assigned to {Alice, Carol, Edwina} directly on the basis of a given body of evidence, but also the masses so assigned to {Alice} and all other subsets of the set of women. This is because if Alice did it, then a woman did it: so direct evidence for Alice increases the overall degree of belief that the perpetrator was a woman. If we think of subsets of the frame of discernment as propositions then the idea is that the overall degree of belief in A is the sum of all the masses assigned to propositions that imply A: propositions such that if any of them is true, then A is true.

Now what happens when we want to combine evidence from different sources—say from independent witnesses? Different mass assignments, induced by distinct bodies of evidence, are combined by Dempster's rule of combination. Suppose that mass assignment m_1 has focal sets $A_1, \ldots, A_k$ and mass assignment m_2 has focal sets $B_1, \ldots, B_l$ (and—this will be explained shortly—suppose $\sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) < 1$). Then their combination is the mass assignment m defined by $m(\emptyset) = 0$ and for all nonempty $A \subseteq \Omega$:

$$m(A) = \frac{\sum_{i,j:A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)} \tag{4}$$

The core idea here is that the combination of the evidence that induces m_1 and the evidence that induces m_2 warrants assigning mass to a set A if m_1 and m_2 assign mass to sets whose intersection is A. Specifically, it warrants assigning the product of the masses assigned to these sets by m_1 and m_2. Of course there may be more than one pair of focal sets of m_1 and m_2 whose intersection is A, so we sum across all such pairs. That gives the numerator of (4). The reason for the inclusion of the denominator is that there may be pairs of focal sets of m_1 and m_2 whose intersection is the empty set. But one of the defining conditions of a basic mass assignment (1 in (1)) is that the empty set be assigned 0. So we 'confiscate' this mass from the empty set and then—in order that the assignments made by m should sum to 1 (this being the other defining condition of a basic mass assignment: 2 in (1))—add it to the assignments to non-empty sets. Specifically, we renormalise by dividing each such assignment by 1 minus the total amount that would have gone to the empty set—the denominator of (4)—thus proportionally inflating each assignment so that together they sum to 1. The reason for the parenthetical condition above can now be seen. If it is not met, then (4)

---

[11] What number p is will depend on the details of the case (how trustworthy the witness is taken to be, how good a view she had, etc.); I discuss this further in other work. For discussion of the idea that $1 - p$ is assigned to Ω rather than to {Bob, Dave, Frank}, see Sect. 5.

is mathematically undefined (it involves division by 0). In this case m_1 and m_2 flatly contradict each other and cannot be combined.[12]

We have seen how a basic mass assignment generates a belief function via Eq. (2). The notion of a belief function can also be defined abstractly (i.e. conditions given such that any function that meets them is called a 'belief function') and any such function *Bel* can then be shown to induce a function m that satisfies the defining conditions in (1) of a basic mass assignment (which in turn generates *Bel* again via (2)). Talk in terms of basic mass assignments and talk in terms of belief functions is thus intertranslatable.[13] For example, I presented Dempster's rule of combination as a means for combining basic mass assignments but it can also be presented in a translated form as a means for combining belief functions. Nevertheless, belief functions that come from nowhere are of little interest to us here. The aim is to show that belief functions provide a better account than imprecise probabilities of how degrees of belief can be appropriately responsive to evidence. Accordingly, our chief concern is with belief functions that come from a mass function m which is induced by a body of evidence.

To conclude this section, consider two ways of characterising belief functions that, while not incorrect, fail to get to the heart of the matter. First, it is common to characterise the difference between probability functions and belief functions in terms of additivity: probability functions are additive and (in general) belief functions are not. Of course this difference is a genuine one—but it is simply a symptom of the really crucial difference. Both the probability model and the belief function model of degrees of belief share a common starting point. Evidence can warrant greater or lesser degrees of belief in different propositions. We model this with the idea of a unit mass of belief which is to be distributed among propositions (or subsets of the sample space) with the mass assigned to P representing the degree to which the evidence supports P (specifically). We then calculate final credence values for sets by summing the masses assigned to their subsets. The *crucial* difference is that in the probability model, the initial mass must be fully distributed amongst singleton subsets—it is part of the setup that the force of the evidence must ultimately bear on maximally specific propositions—whereas in the belief function model this need not be the case. This then *generates* additivity of final values in the probability model, but not in the belief function case. In both models, belief mass accumulates up the food chain, so to speak, as we ascend from subsets to supersets. The crucial difference is that in the belief function model, mass can be attached *directly* to higher-level sets—thus generating super-additive behaviour—whereas in the probability model, basic masses must all be attached at the bottom level. So additivity is merely a symptom: the heart of the matter is whether evidence can warrant attaching some mass of belief *directly* to a non-singleton subset of outcomes.[14]

Second, note that there is an injective mapping from belief functions to sets of probabilities. A belief function *Bel* generated (via Eq. (2)) by a basic mass assignment m is mapped to the set of Bayesian belief functions (i.e. probability functions) where each Bel' in the set is generated by a mass assignment m' that arises from m by

---

[12] But see n.9. In Smets's treatment, the denominator of (4) is omitted (Smets, 1988, §3.3 and p. 267).

[13] See Shafer (1976, ch. 2–3) for details.

[14] Cf. Sect. 7.

dividing basic masses assigned by m to non-singleton focal sets, amongst the singleton subsets of those focal sets.[15] The lower envelope of this set of probability functions (i.e. the function that assigns to each subset S of the sample space, the infimum of the assignments made to S by the probability functions in the set) is then *Bel* itself (Halpern, 2017, p. 37). This means that it is not incorrect to think of a belief function as a lower bound on a set of probabilities.[16] Nevertheless this should not be our sole or even primary way of thinking about belief functions.[17] For thinking this way makes belief functions seem derivative from and more complex than probability functions: we go from one probability function, to a set of them, to a belief function as lower bound. However there is a clear sense in which belief functions are simpler and more fundamental than probability functions.[18] A probability function is determined by an assignment of a unit of mass subject to the constraint that positive mass may be assigned only to singleton subsets of the sample space Ω. A belief function is determined by an assignment that removes this constraint and is subject only to the minimal requirement that $\emptyset$ be assigned 0. We may therefore regard a belief function as a more or less *minimal* formal representation of the very idea of (numerical) degree of belief: an assignment of numbers to propositions or subsets—directly modelling strength of belief in them—determined by an assignment of a unit of belief mass subject only to more or less minimal structural conditions.[19] This idea is conceptually prior to—not derivative from—the idea of a probability function. We get from belief functions to probability functions by adding constraints and (in that sense) increasing complexity.

---

[15] In more detail, suppose the frame of discernment Ω is $\{S_1, \ldots, S_n\}$ and the focal sets of m are $\mathbf{S}_1, \ldots, \mathbf{S}_k$ (recall that each focal set is a subset of Ω). We want to consider ways of taking the assignments of masses made by m to focal sets, and dividing them up amongst the *singleton* subsets of those focal sets. Given such a division, let $m_{p,j}$ be the part of $m(\mathbf{S}_p)$ assigned to $\{S_j\}$. We define this generally, for all $1 \leq p \leq k$ and $1 \leq j \leq n$, but because the $m_{p,j}$ arise from the $m(\mathbf{S}_p)$ by dividing the latter among singleton subsets of $\mathbf{S}_p$, we have $m_{p,j} = 0$ if $\{S_j\} \nsubseteq \mathbf{S}_p$, and we have $\sum_{j=1}^{n} m_{p,j} = m(\mathbf{S}_p)$. Now let $m'(\{S_j\}) = \sum_{p=1}^{k} m_{p,j}$. Intuitively this is the sum of $\{S_j\}$'s inheritances from all the divisions of masses assigned to focal sets of which $\{S_j\}$ is a subset. m' is then a basic mass assignment and it assigns positive mass only to singleton subsets of the sample space: so it generates (via Eq. (2)) a Bayesian belief function, i.e. a probability function. Thus each way of dividing up m's assignments yields a probability function. The original belief function (generated by m) is then mapped to the set of *all* probability functions that can be obtained in this way. Note that the same set of probability functions can also be arrived at by a different route: it is the set of all probability functions P such that for all subsets S of the sample space, $P(S) \geq Bel(S)$; see Denoeux (2019, p. 94).

[16] Walley (1991, p. 11, p. 47) e.g. sees them this way. Indeed doing so might be useful in some contexts (e.g. it provides an entry point for mathematical probability theory: see Sect. 6).

[17] Cf. Shafer (1976, ix): "This essay…offers a reinterpretation of Dempster's work…that identifies his "lower probabilities" as epistemic probabilities or degrees of belief…and abandons the idea that they arise as lower bounds over classes of Bayesian probabilities." Also Walley (1996, p. 10): "Every belief function can be represented as a lower envelope of a set of probability measures. This is merely a mathematical representation, however; it is misleading and unnecessary to regard a belief function as a lower bound for some unknown probability measure." And Halpern (2017, p. 37): "while belief functions can be understood (to some extent) in terms of lower probability, this is not the only way of understanding them." Cf. also Wasserman (1990, p. 455).

[18] See Sect. 3.

[19] There do exist notions subject to even more minimal conditions than belief functions, e.g. the plausibility measures of Halpern (2017, §2.10), of which belief functions and probability functions are instances.

3 Simplicity

The traditional model of degrees of belief represents them as probabilities. A powerful and widespread line of objection to this model holds that precise probabilistic degrees of belief cannot, in general, be appropriately responsive to evidence:

> it is sometimes rational to make no determinate probability judgment…refusal to make a determinate probability judgment…may derive from a very clear and cool judgment that on the basis of the available evidence, making a numerically determinate judgment would be unwarranted and arbitrary. (Levi, 1985, p. 396) (cf. also Levi, 1974, pp. 394–395)
>
> The moral would seem to be that if we want to give evidence its due—if we want to maintain that a rational investigator ought to adopt only those states of opinion she has good reason to adopt—we had better conclude that Immodest Connectedness is not a principle you should want to satisfy. (Kaplan, 1996, pp. 27–28) (cf. also p. 24, p. 29)
>
> precise degrees of belief are the wrong response to the sorts of evidence that we typically receive…since the data we receive is often incomplete, imprecise or equivocal, the epistemically *right* response is often to have opinions that are similarly incomplete, imprecise or equivocal. …Precise credences…always commit a believer to extremely definite beliefs…even when the evidence comes nowhere close to warranting such beliefs (Joyce, 2010, p. 283, p. 285)
>
> When there is little or no relevant evidence, the probability model should be highly imprecise or vacuous. More generally, the precision of probability models should match the amount of information on which they are based…'Vagueness' or lack of information should be reflected in imprecise probabilities (Walley, 1991, p. 34, p. 246) (cf. also p. 7)
>
> When evidence is essentially sharp, it warrants a sharp or exact attitude; when evidence is essentially fuzzy—as it is most of the time—it warrants at best a fuzzy attitude. (Sturgeon, 2008, p. 159)

The kind of evidential situation that Walley is talking about when he speaks of there being "little or no relevant evidence" has been widely thought to show that we must replace precise probabilities with imprecise probabilities. I agree that it poses a severe problem for probabilism—but shall argue that if we want a formal model of evidence-respecting degrees of belief then we have better reasons to employ belief functions than imprecise probabilities.

Let's start with the problem for the traditional view. Suppose that Max has absolutely no evidence about the bias/fairness of a given coin. Given probabilism, he must assign a degree of belief to the proposition that the coin will land heads that is representable as a real number between 0 and 1. Different versions of precise probabilism will make different recommendations—for example, Max might apply the principle of insufficient reason and assign 0.5, or he might follow a whim and assign (say) 0.3—but the point to be made below goes through in any case. Now suppose Max has the coin examined by a laboratory which reports that it is balanced in such a way as to

give a chance of 0.5 (or 0.3) of heads. Max's degree of belief remains 0.5 (or 0.3).[20] But then that belief state is not appropriately responsive to the evidence. In one case Max has no evidence but assigns degree of belief 0.5 (or 0.3); in the other case he has lots of evidence but still assigns degree of belief 0.5 (or 0.3).

By contrast, the situation is nicely handled by the belief function model. Taking the frame of discernment to be $\{H, T\}$, where H (T) is the situation in which the coin comes up heads (tails), the case of total lack of evidence is modelled by a basic mass assignment that assigns 1 to $\{H, T\}$. The case of strong evidence for a 0.5 (or 0.3) chance of heads is modelled by a basic mass assignment that assigns 0.5 each to $\{H\}$ and $\{T\}$ (or 0.3 to $\{H\}$ and 0.7 to $\{T\}$—either way, this assignment generates a Bayesian belief function). As desired, the agent's belief state when he has no evidence is different from his belief state when he has definitive evidence of a specific bias. Furthermore, the belief states are not simply different: the second arises from the first in light of the new evidence. Max begins with no evidence and hence assigns his entire belief mass to the whole frame of discernment (for he has no evidence on the basis of which to make any more specific assignment). He then gets evidence about the coin and on its basis he reassigns his belief mass accordingly.[21] Thus the model captures the idea of degrees of belief that respond to or respect the evidence.[22]

What about the imprecise probability model? Like the belief function model— and unlike the precise probability model—it offers different representations of Max's belief states when he has no evidence and when he has definitive evidence of a specific bias. As in the belief function model, the case of strong evidence for a 0.5 (or 0.3) chance of heads is modelled by a single probability function that assigns 0.5 each to $\{H\}$ and $\{T\}$ (or 0.3 to $\{H\}$ and 0.7 to $\{T\}$). The case of total lack of evidence is modelled differently: by the set of *all* probability functions over the sample space.

The argument for the imprecise probability model over the precise probability model is that the former allows us to distinguish cases that the latter forces us to conflate.[23] My argument for the belief function model over the imprecise probability model does not—indeed cannot—take the same form. The imprecise probability model can always distinguish wherever the belief function model can, because there is an injective mapping from belief functions to sets of probabilities (see Sect. 2). However this still leaves room to argue that the belief function model is superior to the imprecise probability model in other ways. In general, there is an abundance of well-studied formal structures, together with well-known methods for encoding the

---

[20] Cf. Popper's (2002, pp. 425–426) 'paradox of ideal evidence'.

[21] In more detail, there are actually two subtly different ways of viewing the reassignment process. First, one could say that Max's initial complete lack of evidence warrants a basic mass assignment m_1 of 1 to the frame of discernment $\{H, T\}$. The evidence from the laboratory warrants a basic mass assignment m_2 of 0.5 each to $\{H\}$ and $\{T\}$ (or 0.3 to $\{H\}$ and 0.7 to $\{T\}$). Combining m_1 and m_2 using Dempster's rule (see Sect. 2) gives a basic mass assignment m_3 that, like m_2, assigns 0.5 each to $\{H\}$ and $\{T\}$ (or 0.3 to $\{H\}$ and 0.7 to $\{T\}$). The other way of looking at the situation is that Dempster's rule does not come into play because the rule tells us how to combine evidence, and Max initially has no evidence at all: so when evidence emerges, m_1 is simply replaced by m_2 (rather than being combined with it). Whichever interpretation we place on the situation, the net result is the same because m_2 and m_3 make the same assignments.

[22] See also Shafer (1976, pp. 22–25) for a somewhat different argument that the belief function view models ignorance better than the precise probability view.

[23] Cf. e.g. White (2010, p. 173), Bradley (2019, §2.4) and Genin and Huber (2021, §3.15).

objects in certain structures into those in others. Given a formal model that can mark the distinctions we regard as important, we can get another formal model that is also adequate in this respect by encoding the objects used in the first modelling system into a different bunch of objects.[24] Yet we do not regard all such models equally favourably. Having the resources to mark intuitively important distinctions (such as that between having no evidence, and having strong evidence for no bias) is a necessary condition of adequacy on a formal model but there are also other desiderata. One of them is simplicity—or at least, lack of needless complexity. This is the first respect in which the belief function model shows its superiority.

Consider the situation where the agent has no evidence at all. Having absolutely no evidence is an incredibly simple kind of situation to be in. The belief function model represents it in a correspondingly simple way: an assignment of 1 to the frame of discernment. The imprecise probability model, by contrast, represents it in a highly complex way: an infinite set containing every possible probability function over the sample space. This is not only complex but also indirect: when one wades through this infinite set of precise assignments of this or that value to this or that subset of the sample space, one finds that the information it is supposed to encode is, essentially, nothing. The representation is like a politician's speech: an overabundance of words masking a lack of content. The problem is not the lack of content: the point is to represent an 'empty' belief state—the result of a total lack of evidence. Rather, the complaint concerns the indirection and complexity of the representation. If you have nothing to present, better to do so straight up—as in the belief function representation—than wrapped in infinite layers like a giant pass-the-parcel with no central prize.

The core of the issue here is that the imprecise probability view has no *simple* or *direct* way of representing *lack* of evidence for P. This becomes a problem when the task is precisely to represent the belief states of agents with a paucity of evidence. In the belief function framework, having no evidence for P is represented by assigning 0 to P. This is as simple and direct a representation as one could hope for: no evidence, zero degree of belief. The guiding idea—the design brief—was to model *evidence-respecting* or *evidence-responsive* degrees of belief. And in this model, if there is no evidential pressure on the accelerator pedal, then indeed the engine of belief is not revving. Or to vary the image, picture evidence-responsive degrees of belief, as modelled by belief functions, in terms of an evidenceometer: strong evidence deflects the belief needle significantly, up towards 1; weak evidence deflects it a little bit, not so far above 0; and if there is *no* evidence then the needle does not move at all and just sits on 0. But this idea—that having no evidence for P is represented by assigning 0 degree of belief to P—does not work in the probabilistic framework, because 0 probability for P goes hand in hand with probability 1 for $\neg P$. So if you have no evidence at all (for P or $\neg P$) then you cannot assign 0 to either of them (for then you would have to assign 1 to the other). So the imprecise probabilist represents a lack of evidence for P by an infinite set of probability functions which between them make all possible assignments to P. And while this captures the distinction between lack of

---

[24] To take a familiar example, a model of propositions as formulas in a certain formal language can be turned into a model of propositions as numbers by Gödel numbering the formulas. Any distinction between propositions that can be marked in the first system by modelling the propositions as distinct formulas can also be marked in the second system by modelling the propositions as distinct numbers.

evidence for P and $\neg P$, versus evidence that they are equally likely, it is nevertheless a roundabout and complex representation of the former situation—compared to the simple and direct representation in the belief function framework.[25]

The situation is similar when there is some lack, but not a total lack of evidence. Consider Ellsberg's (1961, p. 653) "urn known to contain 30 red balls and 60 black and yellow balls, the latter in unknown proportion." Here we have evidence regarding the number of black-or-yellow balls but no evidence regarding the number of black (specifically) or yellow (specifically) balls.[26] The belief function representation is simple and direct. Let the frame of discernment be $\{R, B, Y\}$, where $R/B/Y$ is the situation in which a red/black/yellow ball is drawn. The basic mass assignment assigns $\frac{1}{3}$ to $\{R\}$ and $\frac{2}{3}$ to $\{B, Y\}$. So we have zero degree of belief in $\{B\}$ and in $\{Y\}$: zero evidence for black (specifically) so zero degree of belief (and similarly for yellow). As before, the imprecise probabilist cannot take this simple route. She represents the situation as an infinite set of probabilities, each of which assigns $\frac{1}{3}$ to $\{R\}$, and numbers x and y to $\{B\}$ and $\{Y\}$ such that $x + y = \frac{2}{3}$. The imprecise probability framework can certainly distinguish between the Ellsberg case and the case in which we know that there are (say) 40 black balls and 20 yellow ones—but its representation of the former situation is more complex and less direct than the belief function representation. The latter is totally transparent: you have evidence regarding the number of black-or-yellow balls so you assign $\{B\} \cup \{Y\}$ a positive degree of belief; you have no evidence regarding the number of black (specifically) or yellow (specifically) balls so you assign 0 to $\{B\}$ and 0 to $\{Y\}$.

When one compares the route from precise probabilism to belief functions with the route from precise probabilism to imprecise probabilism, it isn't surprising that imprecise probability models end up being more complex and less transparent. The

---

[25] A referee raised the following objection to the idea that in the belief function framework, having no evidence for P is represented by assigning 0 to P. Suppose you start with no evidence at all about a certain coin and so assign 0 to the proposition that it will come up tails on the midday toss. The coin is tossed at midday and you observe that it comes up heads. Now you still assign 0 to the proposition that the coin lands tails on the midday toss. So your degrees of belief have *not* responded to the acquisition of excellent—indeed completely decisive—evidence. Hence if the belief function view models complete lack of evidence in terms of assignments of 0, it faces the same kind of problem as the one noted above for traditional precise probabilism. In response, I note that in the belief function framework, one has to decouple evidence for P and evidence for $\neg P$ (more about this in Sect. 5). In the example just given, you start with no evidence at all for the proposition that the coin will land tails on the midday toss, and so (I say) you assign it 0 degree of belief. After seeing the coin land heads you *still* have no evidence at all for the proposition that the coin landed tails and so you should indeed still assign it 0 degree of belief. You did not get any evidence *that the coin landed tails*—you have no more evidence for that proposition than you had before—and so your degree of belief in that proposition should not change from 0. What you got evidence for was the proposition that the coin landed heads. Indeed you got conclusive evidence and so your degree of belief in that proposition should go to 1. So this example does not cast doubt on the claim that the belief function model accommodates the idea that degrees of belief should respond to evidence. What the example does is underline the fact that the belief function model decouples evidence for and hence degrees of belief in P, and evidence for and hence degrees of belief in $\neg P$. By contrast, in probabilistic models, degrees of belief in P and in $\neg P$ are tightly linked. This issue will be discussed further in Sect. 5.

[26] Of course we know that there cannot be more than 60 black balls (and we know that there cannot be more than 60 yellow balls). But this isn't based on evidence about the number of black balls *specifically* (or about the number of yellow balls *specifically*). It is based on evidence about the number of black-or-yellow balls. Because we know this number is 60, we can infer that there cannot be more than 60 black balls (or more than 60 yellow balls).

imprecise probability strategy for moving away from precise probabilities takes something that does not, in general, provide a good model of evidence-responsive belief states—a probability function—and *multiplies* it, to give a (generally infinite) *set* of such things. The belief function strategy takes the probability function and *simplifies* it, by removing one constraint from its definition. Instead of requiring basic belief masses to be distributed amongst *singleton* subsets of the sample space, they may be distributed amongst *any* subsets.[27] The imprecise probability model multiples a constrained thing. The belief function model stays with a single thing but removes a constraint. The former moves in the direction of increasing complexity. The latter moves in the opposite direction.

4 Interpreting the models

While simplicity is desirable in a formal model, another feature is essential: there must be a clear and coherent story about how the models relate to the phenomena. In other work, I argue that the imprecise probability model of degrees of belief lacks such a story. That is, the model lacks a workable interpretation: a viable account of how we are to extract information about an agent's belief state from a set of probabilities. I shall not repeat those arguments here—but I do need to show that the belief function model does have a workable interpretation.

In the traditional probabilistic model of degrees of belief, the interpretative story is simple and direct. A probability function assigns numbers in the interval [0, 1] to propositions, or subsets of the sample space—and those numbers directly represent strength of belief. The belief function model inherits this story. It too assigns numbers to propositions, or subsets of the sample space—and those numbers directly represent strength of belief. 0 means the weakest possible strength of belief. 1 means maximum strength belief. In between, bigger numbers mean stronger degrees of belief and smaller numbers mean weaker degrees of belief. Indeed the story about how the belief function model relates to the phenomena modelled is even more simple and direct than the story about the traditional probabilistic model, for the reason touched on in Sect. 3. In the belief function model, the number assigned to P represents the degree of belief in P that is warranted by the agent's evidence: from 0 (no support) to 1 (complete support). In the probabilistic case the account is not quite so simple because of the bleed-through between the stories about P and $\neg P$. Assigning 0 to P does not simply indicate that the evidence provides no support for P: it goes hand in hand with assigning 1 to $\neg P$ and hence indicates that the evidence warrants such a degree of belief.

---

[27] There is a sense in which removing this constraint increases complexity. Where a probability function is determined by a division of unit belief mass among n elements of the sample space, a belief function is determined by a division of unit belief mass among $2^n - 1$ nonempty subsets of the sample space. But I have already addressed this in Sect. 2. In practice, the belief function will often be determined by a handful of assignments. In the case of total ignorance, the single assignment of 1 to the sample space determines that all other assignments must be 0. In the Ellsberg case, the two assignments of $\frac{1}{3}$ to $\{R\}$ and $\frac{2}{3}$ to $\{B, Y\}$ determine that all other assignments must be 0. And in the worst case scenario, the move from n to $2^n - 1$ assignments (i.e. from probability to belief function) pales in comparison to the move from n assignments to an infinite set of functions each of which makes n assignments (i.e. from probability to imprecise probability).

For certain sets of probability functions, their lower envelope is a belief function.[28] So an imprecise probabilist could, in theory, respond as follows: "I only countenance sets of probability functions whose lower envelopes are belief functions, and the only thing about such a set that I take to be meaningful—to represent something about degrees of belief—is its lower envelope. As for my interpretation of this lower envelope: it is exactly the same as your interpretation of a belief function, which I simply co-opt." But then this imprecise probabilist would be better off jettisoning the set of probabilities—which by her own lights is meaningless baggage—and just taking the lower envelope as her model of degrees of belief. This would simplify her model, without removing anything she regards as important—as conveying information about degrees of belief. But of course it would also render her a belief function theorist.

5 Asymmetric evidence

Consider a detective who starts out with no evidence at all and then gains some evidence for (say) the proposition W that a woman did it. Her degree of belief in W will go up—but what happens to her degree of belief in the proposition M that a man did it? The belief function model has the capacity to say two different things here, depending on the nature of the evidence. First, suppose the situation is as follows. The detective has a bugging device, and she hears a woman's voice on it—but there is a programming error in the bugging system, so that the central recorder is picking up signals from multiple bugs without logging which is which. So there is only a 10% chance that the recording of a woman's voice comes from the scene of the crime (rather than one of the other nine locations she has bugged).[29] In the belief function framework this situation can be modelled as follows.[30] The detective starts with a basic mass assignment m_1 with $m_1(\Omega) = 1$, and ends with a basic mass assignment m_2 with $m_2(W) = 0.1$ and $m_2(\Omega) = 0.9$. Second, suppose the situation is different: there is no bugging device. Instead, the detective learns that a crime syndicate chose which of its operatives would do the job by flipping a coin: heads a woman, tails a man. The detective also learns that the coin used was biased, with a 10% chance of landing heads. In the belief function framework this situation can be modelled as follows. The detective again starts with a basic mass assignment m_1 with $m_1(\Omega) = 1$. This time she ends with a basic mass assignment m_3 with $m_3(W) = 0.1$ and $m_3(M) = 0.9$.

In both cases the detective moves from having zero degree of belief in W to having degree of belief 0.1. The difference between the two cases lies in the treatment of the remaining 0.9 belief mass. In the first case m_2 assigns it to the sample space; in the second case m_3 assigns it to M. Thus in the first case the detective's degree of belief in M does not change: it remains at zero. In the second case the detective's degree

---

[28] See Sect. 2.

[29] This case may be a little far-fetched, but it has the same structure as the much more realistic case of unreliable testimony. The bug is like a witness who says that a woman did it, where the detective—because she thinks the witness may be mistaken or lying or both—gives low credence to the witness telling the truth.

[30] Ω is the frame of discernment, comprising all the possible perpetrators. W (M) is the subset of Ω containing all the women (men). For the sake of example, suppose that in this case $W \cap M = \emptyset$ and $W \cup M = \Omega$.

of belief in M goes up: from 0 to 0.9. The rationale for the different treatment of the two cases is as follows. In the first case, if the evidence ends up being good (i.e. the bug was from the relevant crime scene) then a woman did it. But if the evidence ends up being bad, that does not mean a man did it. It means the detective is back to square one, with no evidence. So the 90% chance that the evidence is bad does not contribute a weight of 0.9 to M: it remains at the level of the sample space. This aspect of belief functions can take a bit of getting used to, for those used to thinking in terms of probability functions—so here's a way of thinking that might be helpful. The notions of evidence and proof are related: proof is the maximum level that evidence can attain. It is a familiar fact (and a significant one in, for example, the context of intuitionist logic) that absence of a proof of P is not necessarily the same as a proof of $\neg P$. This structure carries over to certain kinds of evidence, as modelled by belief functions. When you have (say) 0.7 degree of belief that P, based on evidence E, this does not automatically mean that the 0.3 degree to which E falls short of conclusively supporting P gets applied to $\neg P$: not being evidence for P is not necessarily the same as being evidence for $\neg P$. It can be that the 0.3 goes neither towards P nor towards $\neg P$, but gets applied to the background space of possibilities. In the second version of the detective case however, m_3 does assign the remaining 0.9 to M, rather than to the sample space. This is because the evidence is different in this case. The coin had a 10% chance of landing heads. If it landed heads, a woman did it.[31] But furthermore, if it did not land heads, it landed tails. There is a 90% chance of that. And if that happened, a man did it. So 0.1 goes to W and 0.9 to M.

The imprecise probability framework, on the other hand, cannot model the first kind of response to the evidence: an asymmetric response in which there is a change in the degree of belief in some proposition P (in this case W) but no change in the degree of belief in $\neg P$ (in this case M).[32] In the imprecise probability framework there will always be a symmetry between the treatments of P and $\neg P$, inherited from the underlying fact about probability functions that the probabilities of P and of $\neg P$ must sum to 1. If the probability of P is spread between m and n, then the probability of $\neg P$ must be spread between $1 - n$ and $1 - m$. If the minimum value assigned to P by probability functions in the set goes up (down), then the maximum value assigned to $\neg P$ goes down (up). If the maximum value assigned to P goes up (down), then the minimum value assigned to $\neg P$ goes down (up). In general, the treatment of $\neg P$ will always be the mirror image of the treatment of P, reflected around 0.5. The imprecise probabilist could reply that she cares *only* about minimum values (or maximum values—but one or the other, not both). For the minimum value assigned to P could move without the *minimum* value assigned to $\neg P$ moving. However imprecise probabilists tend to be keen to distance themselves even from the view that the only information conveyed by a set of probabilities is the lowest *and* highest values assigned

---

[31] Assume the detective is certain that no syndicate member would fail to follow a directive to commit the crime.

[32] If the belief function theorist models the detectives in the two situations (after the evidence comes in) using two different belief functions, then the imprecise probabilist can of course find two different sets of probabilities with which to model them (see Sect. 3). My point is that the imprecise probabilist's model of the first kind of response to the evidence cannot be characterised as one in which there is a change in the degree of belief in P but no change in the degree of belief in $\neg P$.

to each proposition—let alone the view just mooted—precisely on the grounds that sets of probabilities convey more information.[33] An alternative line of reply for probabilists might be to try to deny the intuition that certain bodies of evidence can *partially* support *P*, without providing *any* support to ¬*P*.[34] But this would be a hard row to hoe, for the intuition is both long-lasting and widespread. Its force has been felt right back to the early days of probability theory, by such luminaries as Leibniz and Jacques Bernoulli.[35] And more recently, it has been expressed not only by historians of those early days[36]—nor only by proponents of belief functions such as Shafer (1976)—but also by proponents of distinct models, for example Cohen (1977) on Baconian (aka inductive) probability.

6 The compatibility argument

I have argued that the belief function framework is superior to the imprecise probability framework in several respects. Now I shall strengthen the relative position of the case for belief functions vis-à-vis the case for imprecise probabilities not by bolstering the former but by undercutting the latter. I identify two main kinds of argument for imprecise probabilities as a model of evidence-respecting degrees of belief. This section argues that the first kind fails. The next section argues that the second provides at least as much support for the belief function view.

The first kind of argument parallels the *plurivaluationist* approach to the semantics of vague discourse.[37] The classical approach to the logical formalisation of a discourse (exemplified for instance in the first-order formalisation of arithmetic) defines a formal language, and associates statements of the discourse with well formed formulas (wffs) of the language. It then captures the idea that some statements in the discourse are simply true by distinguishing, amongst the infinity of interpretations of the formal language—on some of which the wff associated with a given statement is true and on some of which it is false—a unique *intended* interpretation.[38] Truth (simpliciter) of a statement in the discourse becomes (model-theoretic) truth of the associated wff on the intended interpretation. The plurivaluationist approach to the semantics of vague

---

[33] See e.g. Levi (1980, §9.8) and Joyce (2010, pp. 294–296).

[34] Cf. Shafer (1978, p. 332): "The contemporary understanding of probability is so dominated by the model of chance that nonadditive probabilities are apt to strike us as bizarre, artificial, or simply impossible."

[35] Cf. Bernoulli's notion of pure, as opposed to mixed, evidence—and his related conception of non-additive probabilities—and Leibniz's related notion of pure, as opposed to mixed, proofs in law; see Hacking (2006, ch. 16, ch. 10) and Shafer (1978).

[36] See for example Hacking (1974, p. 114) (cf. also Hacking, 2006, pp. 152–153), who gives an everyday example of evidence that intuitively confers 90% certainty on a certain proposition without conferring 10% certainty on its negation.

[37] This approach was traditionally labelled 'supervaluationist' but following Smith (2008, §2.5) it is now often called 'plurivaluationist' to distinguish it from a view that was also traditionally labelled 'supervaluationist' but is in fact importantly (albeit subtly) different. Note that in this paper I am concerned only with *classical* plurivaluationism—which I simply call 'plurivaluationism' here—as distinct from other plurivaluationist views such as the fuzzy plurivaluationism proposed and defended in Smith (2008).

[38] I use the term 'interpretation' here instead of the nowadays more common 'model' because the latter term is already used with a different sense in this paper (in talk about formal models of degrees of belief).

discourse differs from this classical approach in only one respect: in place of a unique intended classical interpretation, it posits multiple acceptable classical interpretations.

There are two different kinds of thought that can be taken to motivate the pluri-valuationist approach: one turning on the notion of *precisification*, the other on the notion of *constraint*. Starting with the former, the thought is that as it stands, ordinary vague language is not well modelled by a classical interpretation. (For example, our usage of the term 'tall' does not divide people neatly into two camps—the tall and the non-tall—whereas a classical interpretation assigns each predicate as extension a crisp subset of the domain, which divides it cleanly into two parts.) However, we could in principle completely precisify our language and then a classical interpretation would be an appropriate model of its semantics. But there are many different admissible ways in which we could precisify the language (e.g. we could set the cutoff for tall at $6'$, or at $6'1''$). So we take as our model of the semantics of vague language the set of all interpretations, each of which would be the unique correct interpretation were we to precisify the language completely in one of the many admissible ways in which this could be done. The latter thought is a bit different. The idea is that usage places *constraints* on the correct interpretation of our language. For example, when we (the speakers) all deem a certain person to be clearly tall, we thereby rule out as incorrect an interpretation in which this person is not in the extension of (the predicate of the formal language associated with the word) 'tall'. Where language is vague, however, these constraints fail to winnow down to a unique intended interpretation. Rather, we are left with a set of equally acceptable interpretations—and we take this set as our model of the semantics of vague language. Note that where the former thought gives a central place to the notion of precisification and to what *would* be a correct interpretation of the language were we to precisify it, the latter thought makes no reference to precisification: it considers interpretations that are compatible with (not ruled out by) our actual usage of vague language.

We can find a similar division of views in the imprecise probability literature. On the precisification side we have Jeffrey:[39]

> In practice, we achieve coherence only for preference rankings that involve small numbers of propositions: for small fragments of what our total preference rankings would be, had we the time, intelligence, experience, sensitivity, and patience to work them out. (Jeffrey, 1965, p. 533)
>
> 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, i.e., they are rankings of various subsets of the full algebras. Then even if my theory were like Savage's in that full rankings of whole algebras always determine unique probability functions, the actual, partial rankings that characterize real people would determine belief states that are infinite sets of probability functions on the full algebras…The role of definite probability measures in probability logic as I see it is the same as the role of maximal consistent sets of sentences in deductive logic…[they] play the role of unattainable completions of consistent human belief states. (Jeffrey, 1983, pp. 139–140)

---

[39] Cf. also Kaplan (1989, pp. 62–63), and Jeffrey (1987, p. 589).

On the constraint side we have Levi, Kaplan, van Fraassen and Joyce:

> My thesis is that X's credal state relative to K—i.e., the set of Q-functions seriously permissible according to X relative to K—should consist of all those Q-functions which X has no warrant for ruling out as impermissible relative to K. (Levi, 1980, p. 89)
>
> giving evidence its due requires that you rule out as too high, or too low, only those values of *con* which the evidence gives you *reason* to consider too high or too low. As for the values of *con* not thus ruled out, you should remain undecided as to which to assign. (Kaplan, 1983, p. 570)
>
> a finite and even small number of judgements may convey all there is to our opinion. But then there is a large class of probability functions which satisfy just those judgements, hence which are compatible with the person's state of opinion. Call that his or her *representor* (*class*). (van Fraassen, 1990, p. 347)
>
> Elements of C are, intuitively, probability functions that the believer takes to be compatible with her total evidence. (Joyce, 2010, p. 288)

Either way, to get an argument for imprecise probabilism we need a further step: an argument, in effect, as to why probability functions should be in the picture at all. Suppose we accept (a) that there are facts about an agent's belief state that definitely rule out modelling it as certain probability functions, (b) that there are multiple probability functions that are not definitely ruled incorrect in this way, but (c) that nevertheless there are reasons against picking a single one of them as the model of the agent's belief state. It simply does not follow that we should take the set of not-definitely-incorrect probability functions as our model of the agent's belief state. For everything we have said is also perfectly compatible with taking a single belief function as our model of the agent's belief state. (The definitely incorrect probability functions would then be seen as the ones that are not compatible with the belief function: i.e. not in the set of probabilities to which the belief function is mapped by the injective mapping discussed in Sect. 2.)[40]

The point becomes very clear if we consider the analogous case of plurivaluation-ism about vagueness. Suppose we accept (a) that there are facts about (the usage of speakers of) the vague discourse that definitely rule out modelling it using certain classical interpretations, (b) that there are multiple classical interpretations that are not definitely ruled incorrect in this way, but (c) that nevertheless there are reasons against picking a single one of them as the model of the semantics of the discourse.

---

[40] Indeed in Levi's case this move would seem to be *more* in keeping with his overall viewpoint. As in the quote just above, and as will be discussed further in Sect. 7, Levi thinks that an agent faced with multiple equally admissible probabilistic belief states should remain suspended between them. But when it comes to acts, he writes "I contend that when two or more feasible options are E-admissible and there is a feasible option which expresses suspension of judgment between them, it should, everything else being equal, be chosen" (Levi, 1980, p. 97). The analogue of the latter view in the case of belief would be to adopt a single belief state represented by a belief function—rather than remaining suspended between many belief states represented by probability functions—because a belief function *can* be seen as expressing suspension of belief between the multiple probability functions compatible with it. (Note that although belief functions can be seen in this way, as discussed in Sect. 2 we should not primarily view them thus.) Indeed note that on the page prior to the above quote, Levi (1980, p. 88) writes: "When X is in a predicament where he lacks a warrant for choosing one credal state rather than another, wisdom dictates that X should suspend judgment between the alternatives—i.e., select a credal state which expresses such suspension of judgment."

(For example, we might agree with an epistemicist along the lines of Williamson (1994) that meaning facts must be determined by usage facts, but disagree that the usage facts suffice to determine a single correct classical interpretation.) It simply does not follow that we should take the set of not-definitely-incorrect classical interpretations as our model of the semantics of the vague discourse. For everything we have said is also perfectly compatible with taking a single non-classical interpretation as our model of the semantics: for example, a supervaluationist interpretation or a fuzzy interpretation.[41] Plurivaluationists need a further *argument* for why we should only countenance classical interpretations—and such arguments can indeed be found in the literature. Consider for example Lewis (1986, p. 212):

> The only intelligible account of vagueness locates it in our thought and language. The reason it's vague where the outback begins is not that there's this thing, the outback, with imprecise borders; rather, there are many things, with different borders, and nobody's been fool enough to try to enforce a choice of one of them as the official referent of the word 'outback'.

There is a (more or less implicit) argument here for the plurivaluationist approach, which models vagueness using a multiplicity of classical interpretations rather than one nonclassical interpretation. The argument is that the world itself is precise: it offers up only precise objects and precise properties. So a model of how we talk about the world *must* be classical: it must represent us as picking out precise objects with our names and precise properties with our predicates. An interpretation in which names pick out unique vague objects and predicates pick out unique but inherently vague (or inherently fuzzy, or inherently partial or gappy) properties is a non-starter. Classical interpretations are the only game in town—so if one such interpretation won't do (i.e. if a purely classical approach to vagueness fails) we'll need to appeal to many of them at once.

I am not endorsing this argument for the plurivaluationist approach to vagueness— but at least it is an argument. In the probability case there is just as much need for an argument as to why probability functions are the only game in town and so if we cannot model our belief states using a unique probability function then we must model them using sets of such functions. For there is a completely different way of generalising the precise probability model: instead of many precise probability functions, we could appeal to a single non-probabilistic function. (Just as, in the vagueness case, instead of appealing to many classical interpretations, we could appeal to a single non-classical interpretation—this is the kind of approach Lewis is ruling out above.) That is what the belief function model does. To motivate the imprecise probability approach, one therefore needs an *argument* as to why one must stick with probability functions. And yet no clear or compelling argument along these lines has been given in the literature. Consider Kaplan (1989, pp. 57–58), who after saying that the set of probabilities model has "undeniable attractions", sets out two of them—and then comes to the third:

---

[41] In both cases the not-definitely-incorrect classical interpretations would be seen as those classical interpretations that are *extensions* of the unique intended non-classical interpretation: see Smith (2008, ch. 2). Note that I am here using 'supervaluationist' for the view mentioned in n.37 that is different from plurivaluationism.

Finally, the proposal still manages to keep faith with its orthodox predecessor: it requires that each degree of confidence assignment in your set of assignments be an orthodox bayesian degree of confidence assignment (that each of your many minds be an orthodox bayesian mind) and that, as a regulative ideal, each of these assignments obey the axioms of the probability calculus.

This is significant because it makes explicit what otherwise seems to be a pervasive but unspoken assumption in the imprecise probability literature: that probability functions should remain in the picture, even when we abandon traditional precise probabilism. (Note for example that Levi (1980) assumes a setup in which the agent eliminates probability functions insofar as this is warranted. He gives no explicit motivation for why probability functions are part of the setup in the first place, with a right to be there *unless* their elimination is warranted. Similar remarks apply to Kaplan (1996).)

Even in Kaplan, however, it is still an assumption that probabilities should be retained: the passage is quite clearly not an argument for retaining them. (The argument is that imprecise probabilism is good *because* it keeps faith with its orthodox predecessor by retaining probability functions.) I can see two ways to turn this into an argument—but both fail. First, one might argue that in theory change, there is a general methodological principle of minimal disruption: if the received view has problems, mutilate it as minimally as you can manage. But even if we accept this, we need to ask ourselves which is the more minimal change: that from a single probability function to a generally *infinite* set of them; or that from a single probability function to a single function which is just like a probability function except that in its definition we remove one occurrence of the word 'singleton'?[42] I'd argue that the latter change is more minimal. Second, one might argue that mathematical probability theory is rich and mature and, again as a methodological point, we do not want to deny ourselves access to this body of theory. But again this does not favour imprecise probabilism over the belief function approach: both still make ample room for the applicability of probability theory. Recall that a belief function generates a set of compatible probability functions: so probability theory gets ample purchase on belief functions via this associated set.[43] Modelling degrees of belief as belief functions in no way cuts us off from the mathematical riches of probability theory. Compare Gödel numbering, which allows us to apply arithmetical techniques to logic. To do this, we do not literally need to identify formulas with numbers—we only need the right kind of mapping between them.

To sum up, there is a line of thought in the imprecise probability literature that says that facts about an agent's belief state are compatible with multiple probability functions (not just one)—and so we should model the belief state by the set of those functions. I have pointed out that this line of argument is analogous to a line of argument in the vagueness literature for a plurivaluationist approach to the semantics of vagueness over a purely classical approach (such as epistemicism). I have then noted

---

[42] See Sect. 2.

[43] Joyce (2010, p. 292) says that "Requiring each $c \in C$ to be a probability forces graded beliefs to have some nice normative properties." This does not provide a compelling argument for the imprecise probability model over the belief function model, because we can still get access to these properties in the belief function model via consideration of the set of compatible probability functions.

that (as consideration of the vagueness case makes clear) this line of argument requires as a lemma that probability functions (or classical interpretations in the vagueness case) are the only game in town: that the final model must employ them in a central role. I cannot, however, find a convincing argument for this lemma in the imprecise probability literature—and the arguments that I can imagine are not convincing. I conclude that this way of motivating imprecise probabilism fails.

7 The matching argument

Turning now to the second line of argument that I discern in the literature: in the quotes from Joyce, Walley and Sturgeon at the beginning of Sect. 3 we find the following thought. Beliefs should match evidence in kind; evidence is sometimes imprecise or equivocal or vague or fuzzy; so in those cases beliefs should be too. Consider also the following remarks from Levi (1980):

> An ideally rational agent should sometimes suspend judgment by embracing a credal state which is indeterminate in the sense that more than one probability measure is considered permissible for the purpose of computing expected values in deliberation. [10]

> X should not rule out any logically permissible Q-functions. He should endorse a credal state which avoids such elimination of Q-functions without warrant by suspending judgment between them. [89]

> one should not reach unwarranted and arbitrary conclusions. It is preferable to remain in suspense. [95]

Kaplan (1989) says that such an agent "can be said to be of more than one mind" [57] and characterises the imprecise probability view as requiring "that each of your many minds be an orthodox bayesian mind" [58].[44] The idea is that some kinds of evidence warrant the agent's making a judgement—being in a single mind—while other kinds of evidence do not. In these cases the agent should suspend judgement, should be in many minds. So here too we have the thought that sometimes the evidence is sharp and warrants a definite belief state—and sometimes it is not and warrants an indeterminate state.

We can agree with the sentiment that beliefs should match evidence in kind. Indeed I take this to be a facet of the more general idea that degrees of belief should respect the evidence. But there is no good argument here for the imprecise probability view. If we make a certain assumption, then the argument is compelling. But the assumption begs the question and if we remove it, then the argument fails to convince. I'll explain this via consideration of the kinds of cases that Walley et al have in mind: the kinds imprecise probabilists see as involving evidence that is imprecise or fuzzy. At the extreme end we have cases where you have no evidence at all. For an example of a less extreme kind of case, suppose a reliable source tells you that a woman did it. ("That's all I am able to tell you in the circumstances," she says: "but believe me, I

---

[44] Recall also Kaplan's (1983, p. 570) comment (quoted in Sect. 6) about remaining undecided.

know about the case"—and let's suppose you do believe her.) Now suppose we make the assumption that probability functions must be involved in the model of belief. Then we do indeed start to see in these cases indeterminacy, vagueness, suspension of judgement, being in many minds and so on. For undoubtedly, in these cases the evidence rules out some belief states characterised by probability functions, and does not rule in a single such belief state. So there is a set of probabilistic belief states that are not ruled out. Now if we think that probabilities must be a central part of the model of belief states, then it will be natural to see here a kind of vagueness or indeterminacy in the evidence. The vagueness or indeterminacy consists exactly in *not* ruling in a single (probabilistic) belief state. And the natural response to it will be to see the appropriate belief state to be in as one characterised by the whole set of not-ruled-out probability functions. Rather than plump for a single (probabilistic) belief state as a traditional subjective or personalist Bayesian would, we should suspend judgement.

Now that story is OK *if* we assume that probability functions must be involved in the model of belief. But this assumption begs the question in a debate between the imprecise probability and belief function views. And if we drop it and look with clear eyes—unclouded by the thought that probability functions have to be used somewhere—then the cases start to look rather different.[45] First, there doesn't on the face of it seem to be anything vague or fuzzy or indeterminate about having no evidence at all. On the contrary, this is perhaps the most simple and determinate evidential situation one can be in. Compare: if you have a middling amount of money then it might be a very difficult question whether to spend it on A or B or C. But if you have absolutely no money at all then your situation is quite simple and definite: you cannot buy anything. Similarly here: the guiding idea was to model *evidence-responsive* degrees of belief. So you should not have elevated degree of belief in a proposition unless there is a reason, based on your evidence, for doing so. Thus when you have no evidence at all, you should not have an elevated degree of belief in any proposition. The belief function model captures this maximally simple evidential situation in a simple and direct way. It represents your belief state by an assignment of 1 to the entire sample space—and hence 0 to every contentful or nontrivial proposition.

The situation is similar if we turn to the less extreme example of the source who tells you that a woman did it. This evidence is not vague or fuzzy: it is simply nonspecific. It quite definitely and determinately lends weight to the proposition that a woman did it, while lending *no* weight to the proposition that Anne *in particular* did it, no weight to the proposition that Betty (specifically) did it, and so on. This evidence simply does not speak to those specific propositions. After all, it would be absurd to go to Anne, or Betty, or any other particular woman, and say: I have evidence that *you* did it! So on the basis of this evidence, you should not be lending elevated weight to these propositions. (Or to vary the example, suppose a parent learns simply that someone at the school has been caught smoking, and when his child returns home, confronts her and says: I have reason to believe *you* were caught smoking today! This is unfair *because* it is

---

[45] Cf. here Stenning and van Lambalgen's (2008) distinction between reasoning *from* an interpretation and reasoning *to* an interpretation.

completely unjustified, on the basis of the available evidence.)[46] The belief function model captures this nicely. On the basis of this evidence, you assign some positive weight *k* to the set of women and 0 to the singleton of each individual woman.

So simply accepting that belief should match evidence in character does not get us to the conclusion that imprecise probabilism is the right model of the kinds of case considered above. Adding the assumption that probability functions must play a central role in the formal model of belief does push us in that direction: but the assumption begs the question. And without it, it is plausible to see the kinds of evidential situation that motivate moving away from precise probabilism as involving not vague or fuzzy evidence, but nonspecific evidence—and to see belief functions as an appropriate model that captures this nicely. The picture is as follows. Sometimes the evidence lends weight to the hypothesis that Anne in particular did it. Sometimes the evidence lends weight to the hypothesis that a woman did it while lending no weight at all to the hypothesis that Anne in particular did it, or to the hypothesis that Betty in particular did it, and so on. In both cases the evidence warrants a definite judgement: there is no vagueness, no suspension of judgement, no being in many minds about whether to attach this weight or that to Anne etc. The difference is simply whether the weight gets attached to less specific propositions or more specific ones. From this point of view, the problem with a probability function when the evidence is nonspecific isn't that it is too *precise*: it's that it is too *constrained*. It forces belief mass to be distributed among singletons in a way that the evidence simply does not warrant. What we need in such cases is not something more vague or fuzzy. We need something less constrained, less specific—and the belief function model provides this.

A probability function cannot capture the idea of evidence-justified degree of belief attaching to less specific propositions without attaching to any more specific propositions that imply them. The probability model forces weight assigned to larger propositions to 'fall through' them, all the way down to their singleton subsets—hence the appropriateness of van Fraassen's muddy diagrams.[47] But opponents of traditional precise probabilism think that sometimes the evidence does not warrant beliefs of this sort. By contrast, a belief function can be nonspecific: it can assign positive weight to a set and zero weight to its subsets. In Sect. 2 I said that additivity is merely a symptom of the core difference between belief functions and probability functions. The heart of the matter is whether evidence can warrant attaching some mass of belief *directly* to a non-singleton subset of outcomes. What the cases of allegedly vague, imprecise and equivocal evidence show is that evidence *can* do this.

It is important to note here that evidence *can* also warrant probabilistic degrees of belief, in some situations: notably, when the evidence consists in observations of *frequencies* or facts about the objective *chances* of outcomes of some randomising device (e.g. a coin, die or roulette wheel). Probability theory is purpose-built to model

---

[46] I intend it to be part of the setup of this case that the parent is not 'putting two and two together'—combining the report from the school with other things seen lately—but is drawing a conclusion *solely* on the basis of the report that someone at the school was caught smoking.

[47] See Sect. 2.

chances and frequencies and it models them well.[48] This is no problem for the belief function view because probability functions are special cases of belief functions. Thus belief functions can model situations in which the evidence does warrant probabilistic degrees of belief, as well as situations in which it does not.[49] Indeed we have here a spectrum. At one end the agent has no evidence at all, and all the weight of belief attaches to the sample space. At the other end the agent has very specific evidence which warrants distributing all the weight of belief among specific outcomes, generating a Bayesian belief function. In between we have cases such as the testimony that the perpetrator was a woman. For each of these kinds of case the belief function framework provides a simple, direct model in which the belief state matches the evidence in its level of specificity.

8 Dilation and inertia

This section concludes the case for belief functions as a model of evidence-respecting degrees of belief by showing that they readily handle two kinds of situation that imprecise probabilities (notoriously) cannot handle.[50]

First, dilation. Consider the following case:

> You haven't a clue as to whether p. But you know that I know whether p. I agree to write 'p' on one side of a fair coin, and '$\neg p$' on the other, *with whichever one is true going on the heads side* (I paint over the coin so that you can't see which sides are *heads* and *tails*). We toss the coin and observe that it happens to land on 'p'. (White, 2010, p. 175)[51]

If we analyse this case using imprecise probabilities, we get an unacceptable result. As you start out with absolutely no evidence for or against p, your belief in p starts out maximally indeterminate: your representor includes probability functions that assign p any value between 0 and 1.[52] You know the coin is fair, however, so these functions all assign 0.5 to h and to t (the propositions that the coin lands with its heads—resp. tails—side up). Now consider what happens when the coin is tossed and you observe that it lands with 'p' showing. Given that 'p' is showing and that the true claim was written on the heads side, out of the original four possibilities ($ph, pt, \bar{p}h, \bar{p}t$) there

---

[48] For example, the requirement that the probabilities of P and $\neg P$ sum to 1 makes perfect sense here. If the proportion of the population—or of the possible outcomes—of which some description is true is $n\%$, then the proportion of which it is not true must be $(1 - n)\%$.

[49] Cf. Sect. 5.

[50] For further support for the belief function model, see Shafer (1976, pp. 25–28) on the fact that Dempster's rule of combination is symmetric (it makes no difference which evidence is old and which is new) and does not require evidence to take the form of a certainty (it accommodates new evidence that warrants only intermediate degrees of belief).

[51] I have changed a 'p' to '$\neg p$' because it is clearly a typo. I use $\neg$ for negation instead of White's $\sim$.

[52] As is standard in the literature, I use 'representor' as a term for the set of probability functions that (according to imprecise probabilism) represents an agent's belief state. This usage derives from van Fraassen: see the quote in Sect. 6. The term is used in a closely related but subtly different sense by Maher (1993, p. 21). A less common term is 'probasition': "Levi (1974) and I take belief states to be sets of probability measures…Call such sets "probasitions." " (Jeffrey, 1983, p. 141).

are now two left open $(ph, \bar{p}t)$.[53] So you conditionalise each probability function in your representor on the set $\{ph, \bar{p}t\}$. The result is that your representor includes probability functions that assign h (i.e. $\{ph, \bar{p}h\}$) any value between 0 and 1: your belief in h has gone from a sharp 0.5 to maximally indeterminate. The same point applies to t. So your beliefs are not respecting the evidence. Nothing happened to undermine your evidence that the coin is fair or to provide further evidence regarding h or t, so your degree of belief in h and t should still be 0.5.

Joyce (2010, 299ff.) argues that the new information *is* evidentially relevant to h. However when followed through closely, his argument consists in showing how this is so *within* the imprecise probability model. That is, Joyce explains in detail how the model leads to this result—not why the result is the right one. The problem is that the information that the coin lands with 'p' showing clearly is not evidence for or against h, by any reasonable pretheoretical reckoning—given that you know that the true one of 'p' and '$\neg p$' is on the heads side *but have absolutely no evidence for or against p*. So Joyce's argument does not avoid the conclusion that the imprecise probability model fails to provide an adequate account of evidence-respecting degrees of belief. Note that *if* you were sure that p (or not p) then seeing 'p' come up would tell you for sure that the coin landed heads (or tails). *If* you thought that p was quite/very un/likely, then seeing 'p' come up would tell you that heads is quite/very un/likely— and so on. But the whole point is that you have no evidence at all about p: not that it is very or a little bit likely or unlikely or anything else. In that situation, seeing the coin land 'p' tells you nothing about whether it landed heads. As we'll see in a moment, this is exactly the result that the belief function story gives. But the imprecise probability story cannot yield this result. Each member of the 'credal committee'[54] cannot avoid taking the initial probability assigned to p as a judgement that p is very (a little bit) un/likely or whatever—depending on whether the probability is very (a little bit) low/high, etc. Hence each committee member cannot avoid taking the observation of 'p' as evidence for or against h.[55] Thus the committee as a whole goes from being clustered at 0.5 on the question of h to scattering across the unit interval.

The belief function model, by contrast, handles the case nicely. Before the toss, your good evidence about the fairness of the coin (and complete lack of evidence about p) warrants a basic mass assignment m_1 that assigns 0.5 to h (i.e. $\{ph, \bar{p}h\}$) and 0.5 to t (i.e. $\{pt, \bar{p}t\}$). Observing the coin land with 'p' showing gives you conclusive evidence (as explained above) that the actual situation is ph or $\bar{p}t$. This evidence warrants a basic mass assignment m_2 that assigns 1 to $\{ph, \bar{p}t\}$. Combining m_1 and m_2 using Dempster's rule (see Sect. 2) gives a basic mass assignment m_3 that assigns 0.5 to $\{ph\}$ and 0.5 to $\{\bar{p}t\}$. Now m_3—like the initial m_1—generates a belief function

---

[53] Note that p in ph (etc.) represents p being the case, i.e. the claim 'p' being true (as opposed to the coin landing with 'p' showing). We could instead use a sample space with eight outcomes of the form $ph'p'$ (representing the situation in which p is the case, the coin lands heads, and the coin lands with 'p' showing), $pt'p'$ etc. This would add some surface complexity but would make no essential difference to what follows.

[54] Imprecise probabilists often use a metaphor according to which each probability function in the representor, taken individually, models a precise probabilistic agent, and together these agents make up a mental committee. See e.g. Joyce (2010, p. 288) and Moss (2018, p. 163).

[55] Even when the initial probability for p is 0.5, the observation of 'p' is still taken as evidence regarding h: just evidence that points equally for and against h.

that assigns 0.5 to h and 0.5 to t. So there is no loss of determinacy in your degrees of belief in heads and tails.

One objection that someone might raise to this treatment of the case concerns the fact that your belief in p after observing the toss is now 0.5 (and likewise for $\neg p$). That is, m_3—unlike the initial m_1—generates a belief function that assigns 0.5 to $\{ph, pt\}$ and 0.5 to $\{\bar{p}h, \bar{p}t\}$. This will seem the wrong result if one thinks that the toss provides no evidence concerning p. And indeed White (2010, p. 176) claims that "Seeing the coin land 'p' should have no affect on your credence in p." But this is incorrect: the toss does provide evidence about p. Consider the person I (i.e. White's "I") who knows whether p and writes on the coin. If I told you 'p' (or '$\neg p$') and you took him to be reliable in his beliefs about p and speaking sincerely, then this would give you strong evidence for p (or $\neg p$). If I wrote the true one of 'p' or '$\neg p$' on *both* sides of the coin (and you knew this), observing the toss land 'p' (or '$\neg p$') would again give you strong evidence for p (or $\neg p$): just as strong as in the direct testimony case (for this is simply a roundabout way of giving testimony). If I wrote the true one of 'p' and '$\neg p$' on one face of a 100-sided die and the false one on all the other faces, observing the roll would again give you strong evidence relevant to p: whatever you see come up ('p' or '$\neg p$') has a 1% chance of being true and so there is a 99% chance that the one you don't see is true. You should then change your belief in p to 0.01 or 0.99 depending on what you see. This is a form of randomised or probabilistic testimony: I tells you whether p in such a way that you have a 1% chance of receiving the true message (and a 99% chance of receiving a false message). Note that I can control the chance of transmission of the true message by selecting a die with a certain number of sides and writing the true message on a certain number of them (and the false message on the others). White's case is simply the one where I writes on one side of a two-sided die (a coin). This certainly provides evidence relevant to p, just like all the other cases of randomised testimony: in this case, evidence that warrants setting your degree of belief in p to 0.5. (Think of it this way: you receive a report that p and there is a 50% chance that it is true.)

Second, inertia. Consider the following case:

> There are two kinds of chemicals used in this lab, X and Y. We know nothing about how common either chemical is or how this test tube came to be filled. But we have no evidence either way bearing on whether it contains X or Y…We should change our opinion in the light of new evidence. Let's run a test that has a 90% accuracy rate (i.e. whether it contains X or Y, the test has a 90% chance of accurately reporting whether it is X or Y). When we get a result, how confident should we be that it is accurate? Obviously 90%. (White, 2010, p. 184)

The imprecise probability view cannot yield this result. Given that an agent knows only that the test tube contains X or Y and has no evidence at all bearing on which one of these two chemicals it is, for every $0 \leq x \leq 1$ her representor should include probability functions that assign x to the proposition that the tube contains X (and $1 - x$ to the proposition that it contains Y). But then, as White and Joyce (2010, pp. 290–291) show, *whatever* evidence the agent gets, her probabilities for X and Y will

remain maximally spread out.[56] This is known as the problem of belief inertia. It is a devastating objection to imprecise probabilities as a model of evidence-responsive degrees of belief. Such degrees of belief should not sharpen in the absence of any evidence—hence they should start maximally spread out in this case—but they should then sharpen when evidence comes in.

One response to this problem on the part of imprecise probabilists is to model the agent's initial state—when she has no evidence—by some more restricted set of probabilities. Joyce (2010, pp. 290–291) outlines a specific proposal along these lines but Vallinder (2018, pp. 1214–1216) shows that it does not achieve the desired result. White also mentions a specific way of restricting the initial set of probabilities but shows that while it does allow the agent's belief state to move in response to the evidence provided by the test result, there is not enough movement: the resulting degree of belief that the tube contains X "is nothing like .9" (White, 2010, p. 185). But there is also a deeper problem with any such strategy of restricting the initial spread of probabilities:[57] it flies in the face of the idea that degrees of belief should respect the evidence. As Joyce (2010, p. 292) says, "what principled reason is there to confine credences to one proper subinterval…as opposed to another? Whichever way you go, you end up acting as if you have evidence that you do not actually possess." As Joyce sees it, "the only view that makes epistemological sense" [290] is the original one according to which the agent's initial belief state is represented by the full, unrestricted set of probability functions. But then we are stuck with the problem of belief inertia, which *also* flies in the face of the idea that degrees of belief should respect the evidence. As White puts it in the quote above, "We should change our opinion in the light of new evidence." So the imprecise probabilist faces a devastating dilemma: either way, she fails to satisfy the original design brief according to which degrees of belief should respect the evidence.

The belief function view, by contrast, handles the situation nicely. Initially, the agent's total lack of evidence about the contents of the tube (apart from knowing that the only possibilities are X and Y) induces a basic mass assignment m_1 that assigns 1 to the frame of discernment $\{T_X, T_Y\}$, where T_X (T_Y) is the situation in which the test tube contains X (Y). Knowing the test to have a 90% accuracy rate and seeing it give the result 'X' (analogous comments apply if it gives 'Y') warrants a basic mass assignment m_2 that assigns 0.9 to $\{T_X\}$ and 0.1 to the frame of discernment. Combining m_1 and m_2 using Dempster's rule gives a basic mass assignment m_3 that is the same as m_2: it assigns 0.9 to $\{T_X\}$ and 0.1 to the frame of discernment.[58] So in light of the test result, the agent's degree of belief that the tube contains X goes from 0 to 0.9, as desired.[59]

---

[56] The same point goes through if we restrict x to the interval (0, 1).

[57] This problem is noted by Joyce and White, and because of it, Joyce does not himself endorse the proposal he outlines for restricting the initial set of probabilities.

[58] Alternatively m_1 could just be replaced by m_2: see n.21.

[59] I am grateful to the anonymous referees for their very helpful comments.

Declarations

References

Bradley, S. (2019). Imprecise probabilities. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring ed.). Metaphysics Research Lab, Stanford University.

Bradley, S., & Steele, K. (2014). Uncertainty, learning, and the "problem" of dilation. *Erkenntnis, 79*(6), 1287–1303.

Cohen, L. J. (1977). *The probable and the provable*. Clarendon Press.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics, 38*(2), 325–339.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B (Methodological), 30,* 205–247.

Denœux, T. (2019). Decision-making with belief functions: A review. *International Journal of Approximate Reasoning, 109,* 87–110.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics, 25,* 643–669.

Genin, K., & Huber, F. (2021). Formal representations of belief. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring ed.). Metaphysics Research Lab, Stanford University.

Hacking, I. (1974). Combined evidence. In S. Stenlund (Ed.), *Logical theory and semantic analysis: Essays dedicated to Stig Kanger on his fiftieth birthday* (pp. 113–123). D. Reidel.

Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference* (2nd ed.). Cambridge University Press.

Halpern, J. Y. (2017). *Reasoning about uncertainty* (2nd ed.). MIT Press.

Jeffrey, R. (1965). Ethics and the logic of decision. *The Journal of Philosophy, 62*(19), 528–539.

Jeffrey, R. (1983). Bayesianism with a human face. In J. Earman (Ed.), *Testing scientific theories, volume X of Minnesota studies in the philosophy of science* (pp. 133–156). University of Minnesota Press.

Jeffrey, R. (1987). Indefinite probability judgment: A reply to Levi. *Philosophy of Science, 54,* 586–591.

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

Kaplan, M. (1983). Decision theory as philosophy. *Philosophy of Science, 50,* 549–577.

Kaplan, M. (1989). Bayesianism without the black box. *Philosophy of Science, 56,* 48–69.

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

Kyburg, H. E., Jr., & Teng, C. M. (2001). *Uncertain inference*. Cambridge University Press.

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

Levi, I. (1980). *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. MIT Press.

Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Philosophy of Science, 52*(3), 390–409.

Lewis, D. (1986). *On the plurality of worlds*. Basil Blackwell.

Maher, P. (1993). *Betting on theories*. Cambridge University Press.

Moss, S. (2018). *Probabilistic knowledge*. Oxford University Press.

Popper, K. (2002). *The logic of scientific discovery*. Routledge.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Shafer, G. (1978). Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences, 19,* 309–370.

Shafer, G. (1979). Allocations of probability. *The Annals of Probability, 7*(5), 827–839.

Smets, P. (1981). The degree of belief in a fuzzy event. *Information Sciences, 25,* 1–19.

Smets, P. (1988). Belief functions. In P. Smets, E. H. Mamdani, D. Dubois, & H. Prade (Eds.), *Non-standard logics for automated reasoning* (pp. 253–277). Academic Press.

Smets, P., & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence, 66,* 191–234.

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B (Methodological), 23*(1), 1–25.

Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford University Press.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science*. MIT Press.

Sturgeon, S. (2008). Reason and the grain of belief. *Noûs, 42*(1), 139–165.

Troffaes, M. C. M., & de Cooman, G. (2014). *Lower previsions*. Wiley.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 2*(42), 230–265.

Vallinder, A. (2018). Imprecise Bayesianism and global belief inertia. *British Journal for the Philosophy of Science, 69,* 1205–1230.

van Fraassen, B. C. (1989). *Laws and symmetry*. Clarendon Press.

van Fraassen, B. C. (1990). Figures in a probability landscape. In J. M. Dunn & A. Gupta (Eds.), *Truth or consequences: Essays in honor of Nuel Belnap* (pp. 345–356). Kluwer Academic Publishers.

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

Walley, P. (1996). Measures of uncertainty in expert systems. *Artificial Intelligence, 83,* 1–58.

Walley, P. (2006). Imprecise probabilities. In S. Kotz, N. Balakrishnan, C. Read, & B. Vidakovic (Eds.), *Encyclopedia of statistical sciences* (2nd ed., Vol. 5, pp. 3353–3356). Wiley.

Wasserman, L. A. (1990). Prior envelopes based on belief functions. *The Annals of Statistics, 18*(1), 454–464.

White, R. (2010). Evidential symmetry and mushy credence. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 3, pp. 161–186). Oxford University Press.

Williamson, T. (1994). *Vagueness*. Routledge.