




Acting on belief functions

Nicholas J. J. Smith¹ 

Accepted: 14 April 2023 / Published online: 13 June 2023
© The Author(s) 2023

Abstract

The degrees of belief of rational agents should be guided by the evidence available to them. This paper takes as a starting point the view—argued elsewhere—that the formal model best able to capture this idea is one that represents degrees of belief using Dempster–Shafer belief functions. However, degrees of belief should not only respect evidence; they also guide decision and action. Whatever formal model of degrees of belief we adopt, we need a decision theory that works with it: that takes as input degrees of belief so represented. The task of this paper is to develop such a decision theory for the belief function model of degrees of belief. This is not the first paper to attempt that task, but compared to the existing literature it takes a more abstract route to its destination, via a consideration of the very idea of rational decision making in light of one’s beliefs and desires. After presenting the new decision theory and comparing it to existing views, the paper goes on to consider diachronic decision situations.

Keywords Degree of belief · Dempster–Shafer belief function · Decision theory · Expected utility · Diachronic Dutch book

1 Introduction

The received model of degrees of belief represents them as probabilities. One of the attractive features of this model is that it integrates with a powerful theory of decision making: expected utility theory. Pretheoretically, degrees of belief feed into behaviour. For example, the greater your degree of belief that Fido is vicious, the wider a berth you will give him. As Ramsey puts it:

the degree of a belief is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it. ...the difference between believing more firmly and believing less firmly...seems to me to lie in how far we should act on these beliefs (Ramsey, 1931, 169, 170)

Expected utility theory gives a formal model of exactly how degrees of belief, represented as probabilities, feed into rational decision and action.

However, there is another side to degrees of belief. Pretheoretically, as well as feeding into behaviour, degrees of belief are also responsive to evidence. For example, as the number of credible, independent witnesses who claim *P* increases, so (other things being equal, and up to a point) should your degree of belief that *P*. This second guiding idea about degrees of belief is found in comments such as:

the degree of support a body of evidence provides for a proposition—i.e., the degree of belief one should accord the proposition on the basis of the evidence (Shafer, 1976, 3)

Epistemic probabilities...depend on the available evidence (Walley, 1991, 14)

One's confidence should adequately reflect one's evidence (White, 2010, 171)

When it comes to this second idea, the model of degrees of belief as probabilities is less attractive. It seems to many authors that frequently the available evidence does not warrant probabilistic degrees of belief. This thought has led these authors to argue that degrees of belief should be formally represented as imprecise probabilities rather than precise probabilities.¹

However, there is little glory in avoiding one problem if one thereby runs headlong into an equally difficult problem—and this is precisely what other authors think the move from precise to imprecise probabilities involves. Specifically, if we model degrees of belief as imprecise probabilities, we need (at least) to modify traditional expected utility theory, or (more radically) to replace it with a quite different account of how degrees of belief feed into rational decision and action—and these authors think that imprecise probabilists have not been able to do this in a satisfactory way.² Thus, while imprecise probabilists may have an advantage when it comes to the link between degrees of belief and evidence, they are at a disadvantage when it comes to the link between degrees of belief and rational action.

The present paper fits into this dialectic as follows. Personally, I am convinced by the arguments that the traditional precise probabilistic model cannot provide a

¹ See e.g. Levi (1985, 396) (cf. also Levi (1974, 394–5)), Kaplan (1996, 27–8 (cf. also 24, 29)), Joyce (2010, 283, 285), Walley (1991, 34, 246 (cf. also 7)) and Sturgeon (2008, 159). 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 probability functions. (See e.g. the opening sentence of Bradley and Steele (2014b): “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.”) 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 probability function. (See e.g. Walley (2006, 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 of these two senses.

² See e.g. Elga (2010), White (2010) and Bradley (2019).

satisfactory account of the link between degrees of belief and evidence, but I am not convinced that imprecise probabilities are the best alternative. In Smith (2022), I present a model of degrees of belief based on Dempster–Shafer belief functions and argue that it provides a superior account of how degrees of belief can respect the evidence.³ This means that I need an account of how degrees of belief, represented as belief functions, feed into rational decision and action. The purpose of the present paper is to provide such an account. So the immediate aim of the paper is to present and argue for a decision theory in which degrees of belief are represented as belief functions. As for the wider significance of the paper, there are two ways of viewing it. First, in accordance with the above account of the surrounding dialectic and my own motivations, one could regard this paper as one part of a larger case for representing degrees of belief using belief functions. Second, one could view the paper as a contribution not to the discussion of how degrees of belief should be represented formally, but to a more abstract discussion of how belief and desire relate to rational decision and action. For my approach will not be to dive straight into the low-level details and try to modify the formalism of expected utility theory so that it can take belief functions (rather than probability functions) as inputs.⁴ Rather, I shall start at a higher level of abstraction, with the very idea of rational decision in light of one’s beliefs and desires. I shall then find an implementation in the setting of belief functions that stands to this abstract idea in the sort of way that expected utility theory stands to the abstract idea in the setting of probability functions. If you think of the abstract idea of decision-making as lying at the apex of a triangle, and expected utility theory at the bottom left corner, then I shall get to a decision theory for belief functions at the bottom right corner, not by going directly across the base, but by going back to the abstract idea of decision making and moving down through the belief function representation of degrees of belief to a decision theory, in a way analogous to the move from this abstract idea, through the probabilistic representation of degrees of belief, to expected utility theory. This relatively abstract approach should be of interest even to those who have no particular sympathy for belief functions. Furthermore, when it comes to comparing the new decision theory to existing views, I shall consider theories that have arisen not only in the literature on belief functions, but also in the literatures on decisions under ambiguity, and on imprecise probabilities.

The paper proceeds as follows. In Sect. 2, I introduce belief functions: not just the formal details but also a way of picturing belief functions that should make them more accessible to intuition. I then turn to the core issue: an account of how the belief function representation of degrees of belief connects up with decision and action. Section 3 presents a first proposal and then shows how it fails. Section 4 presents a better proposal and then Sect. 5 discusses its relationships to existing views. Section 6 replies to a possible objection from a proponent of the traditional model of

³ If, in the present paper, I appear in places simply to assume that degrees of belief need to respect the evidence and that in some situations the evidence does not warrant a probabilistic belief state but does warrant a belief state modelled by a belief function, it should be understood that supporting arguments can be found in the paper just cited.

⁴ Probability functions are special cases of belief functions, so the idea is that the decision theory assumes only that beliefs are represented as belief functions: it works in the case where these belief functions are furthermore probability functions, and in the case where they are not.

degrees of belief as probabilities. Section 7 notes that the proposed link between belief functions and decisions provides criteria for distinguishing different degrees of belief. Section 8 explains how the view of this paper yields a clear and unified conception of the traditional distinction between decisions under risk and decisions under uncertainty. Section 9 replies to two possible objections concerning sequential decisions—including the objection that an agent who makes decisions in the way proposed in Sect. 4 will be subject to diachronic Dutch book.

2 Belief functions

Van Fraassen (1989, 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 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, the new picture involves our apportioning our unit of belief to its non-empty subsets considered individually—that is, the black shaded regions.⁵ 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 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.⁶ 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.⁷

Venn diagrams are useful when the space of possible outcomes is generated by a finite set of basic (logically atomic) propositions (or propositional variables). n such propositions generate 2^n state descriptions: one per assignment of truth values to the

⁵ 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.

⁶ This might seem purely metaphorical—but the idea of mass or weight of belief will be given a more concrete meaning below by the role it plays in decision-making: rational choice involves weighting your desire for an outcome by your degree of belief that an action will lead to that outcome.

⁷ If preferred, you can also vary the image slightly: instead of starting with a rod and cutting it up, you just start by putting whatever weights you like on some or all pans, and then (following van Fraassen) we deem the total of these weights to be one unit.

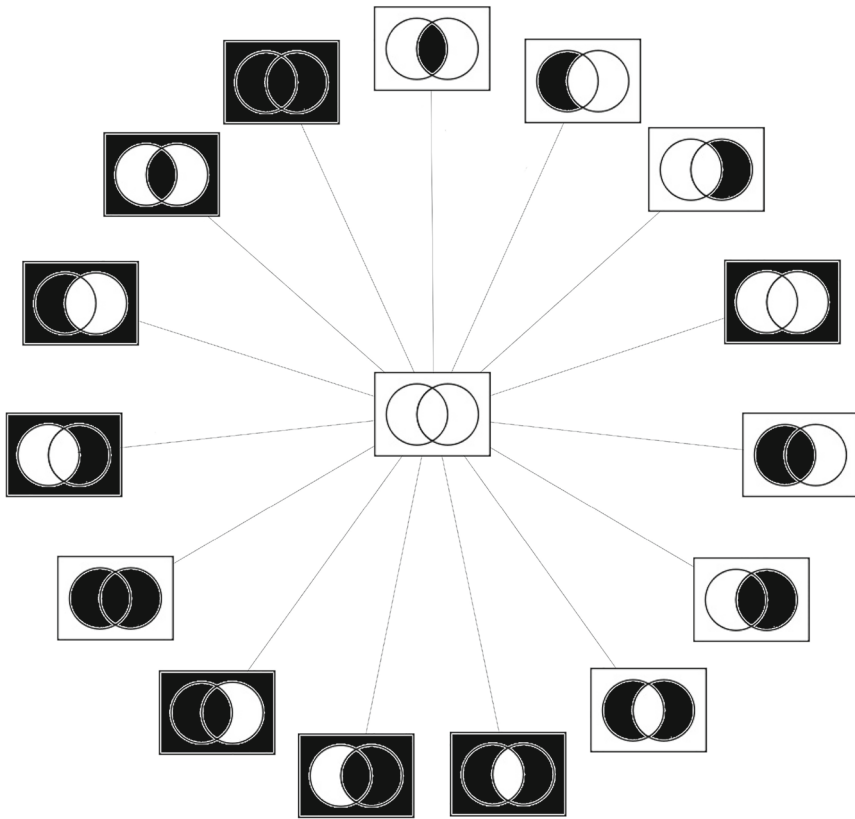


Fig. 1 An exploded Venn diagram

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.⁸ 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 the 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

⁸ E.g. 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).

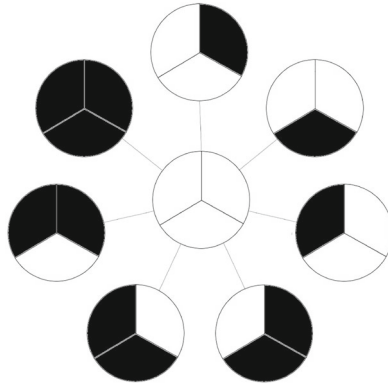


Fig. 2 An exploded non-Venn diagram

weight to each specific outcome: say $\frac{1}{6}$ to each section of the six-section grid.⁹ 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 in the exploded case, one simply places weight a_i on each singleton subset $\{A_i\}$ and places nothing on any other subset.¹⁰ 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. 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)) although it is simpler. But simplicity is not my reason for assuming finitude. Rather, it seems to me 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, Sects. 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 weighting evidence and deciding how to act, the relevant space Ω of possible outcomes will be only

⁹ E.g. each section represents one face of a die and the evidence is that the die is fair.

¹⁰ Cf. the discussion of Bayesian belief functions below.

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).¹¹ Following Shafer (1976, 36), I call Ω the *frame of discernment*.

A *basic mass assignment*,¹² is a function $m : 2^\Omega \rightarrow [0, 1]$ satisfying two conditions¹³:

$$\begin{aligned} 1. & \quad m(\emptyset) = 0 \\ 2. & \quad \sum_{A \subseteq \Omega} m(A) = 1 \end{aligned} \tag{1}$$

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.¹⁴

Given a basic mass assignment m , the corresponding *belief function* $Bel : 2^\Omega \rightarrow [0, 1]$ 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.¹⁵

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:

¹¹ I acknowledge that some readers may find this contentious—and in response, note again that the finitude assumption is not essential.

¹² Terms used in the literature include ‘basic probability assignment’ (Shafer, 1976, 38), ‘basic probability number’ (Smets, 1981), ‘basic belief assignment’ (Smets & Kennes, 1994, 196) and ‘mass function’ (Denoeux, 2019, 93).

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

¹⁴ Condition 1 can be dropped: see Smets (1988). I shall not explore this route here.

¹⁵ It is common to define also the *plausibility function* Pl . It can be defined from m as $Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$ or equivalently from Bel as $Pl(A) = 1 - Bel(\bar{A})$. As Smets and Kennes (1994, 198) write, the plausibility function “is in one-to-one correspondence with belief functions. It is just another way of presenting the same information and could be forgotten, except inasmuch as it provides a convenient alternate representation of our beliefs.” I find it clearer to think in terms of a single kind of attitude—degree of belief or Bel —to different propositions (e.g. A and its negation), rather than two kinds of attitude—degree of belief and degree of plausibility—to a single proposition (A). Hence, plausibility functions play no role in this paper (except as required in the discussion of views of other authors in Sect. 5).

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

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 belief 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 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 masses by equation (2)).¹⁶

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, 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 .¹⁷ 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

¹⁶ 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.

¹⁷ 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.); cf. Sect. 7. For discussion of the idea that $1 - p$ is assigned to Ω rather than to {Bob, Dave, Frank} see Smith (2022, §5).

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, \dots, A_k and mass assignment m_2 has focal sets B_1, \dots, 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 $A \subseteq \Omega$:

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

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) is mathematically undefined (it involves division by 0). In this case, m_1 and m_2 flatly contradict each other and cannot be combined.¹⁸

We have seen how a basic mass assignment generates a belief function via equation (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.¹⁹ For example, I presented Dempster's rule of combination as a way of combining basic mass assignments, but it can also be presented in a translated form as a way of combining belief functions. Nevertheless, belief functions that come from nowhere are of no real interest to us here. As mentioned in Sect. 1, the motivation for considering belief functions is that they provide a good account of how degrees of belief can be responsive to evidence.

¹⁸ But see n.14. In Smets's treatment, the denominator of (4) is omitted (Smets, 1988, Sect. 3.3 and p.267).

¹⁹ See Shafer (1976, ch.2–3) for details.

Accordingly, our chief concern is with belief functions that come from a mass function m that is induced by a body of evidence.

To conclude this section, consider two ways of characterising belief functions that, while not incorrect, in my view 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. This difference is a genuine one—but I see it as 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 ‘mass’ of belief which is to be distributed among propositions (or subsets of the sample space) with the proportion of 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 assignments 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 model. In both frameworks, 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 but 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.

In order to discuss the second characterisation of belief functions, I first need to introduce the following injective mapping from belief functions to sets of probabilities. A belief function Bel generated by a basic mass assignment m is mapped to the set of Bayesian belief functions—i.e. probability functions—where each Bayesian belief function in the set is generated by a mass assignment m' that arises from m by dividing basic masses assigned by m to non-singleton focal sets, amongst the singleton subsets of those focal sets. In more detail, suppose the frame of discernment Ω is $\{S_1, \dots, S_n\}$ and the focal sets of m are S_1, \dots, S_k . Recall that each focal set is a subset of Ω , i.e. each S_p ($1 \leq p \leq k$) is a collection of S_j 's ($1 \leq j \leq n$), i.e. $S_p = \{S_{p_1}, \dots, S_{p_{l_p}}\}$.²⁰ Consider ways of taking the assignments of masses made by m to non-singleton focal sets, and dividing them up amongst the singleton subsets of those focal sets. What we are doing is giving to each singleton subset of S_p a proportion (from 0% to 100%) of the mass originally assigned to S_p (where the proportions add up to 100%). So we can associate each possible way of doing this with a way of proportioning the focal sets. Let a proportioning of a set $S \subseteq \Omega$ be a function $Pr : S \rightarrow [0, 1]$ such that $\sum_{S \in S} Pr(S) = 1$.²¹ Let a proportioning of the focal

²⁰ The reason for having l_p here—rather than just l —is that different S_p 's can contain different numbers of S_j 's.

²¹ $[0, 1]$ is the real unit interval. You can think of its members as percentage values between 0% (i.e. 0) and 100% (i.e. 1).

sets of a basic mass assignment m be a family of proportionings Pr_1, \dots, Pr_k , one for each of the focal sets S_1, \dots, S_k of m . Given a proportioning of the focal sets of m , we can specify a corresponding way of dividing masses assigned by m to non-singleton focal sets amongst their singleton subsets. Let $m_{p,j} = Pr_p(S_j) \cdot m(S_p)$. The idea is that $m_{p,j}$ is the part of $m(S_p)$ that gets assigned to $\{S_j\}$.²² Given this way of dividing masses assigned by m to non-singleton focal sets amongst their singleton subsets, we can then define $m'(\{S_j\}) = \sum_{p=1}^k m_{p,j}$.²³ The idea is that $m'(\{S_j\})$ 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 a Bayesian belief function, i.e. a probability function. Thus each way of dividing up m 's assignments to non-singleton focal sets—each of which corresponds to a way of proportioning the focal sets—yields a probability function. (For future reference, we call these the probability functions that are *compatible* with the belief 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 (i.e. to the set of all probability functions that are compatible with it).²⁴

We can now discuss the second characterisation of belief functions. As we just saw, each belief function Bel can be mapped to a set of probability functions. 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, 37). This means that it is not incorrect to think of a belief function as a lower bound on a set of probabilities.²⁵ Nevertheless this should not be our sole or even primary way of thinking about belief functions.²⁶ For thinking this way makes belief functions seem (conceptually) 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 (conceptually) simpler

²² Note that $m_{p,j}$ is undefined in case $S_j \not\subseteq S_p$ (because Pr_p is defined only on S_p). If you prefer, you could instead define a proportioning of $S \subseteq \Omega$ to be a function $Pr : \Omega \rightarrow [0, 1]$ (i.e. define it on Ω , not on S) and require that $Pr(S) = 0$ if $S \not\subseteq S$.

²³ Here we sum the $m_{p,j}$ that are indeed defined. In case $m_{p,j}$ is undefined for *all* p , we set $m'(\{S_j\}) = 0$. Note that $m_{p,j}$ will be undefined for all p if S_j is not in any focal set of m —but this is not the only way in which $m'(\{S_j\})$ can equal 0: that will also happen if S_j is in one or more focal sets but is assigned zero proportion of each of those sets.

²⁴ 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, 94).

²⁵ Walley (1991, 11, 47) e.g. sees them this way.

²⁶ 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, 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, 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, 455) and Smets and Kennes (1994, 192).

and more fundamental than probability functions. 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. 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 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.²⁷ 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.

3 Degrees of belief and decisions

My goal is to explain how the belief function representation of degrees of belief connects up with decision and action: how belief functions guide the choices of rational agents. This section presents a first proposal and then shows how it fails. Section 4 presents my positive proposal. Section 5 discusses alternative approaches. Before getting to the first proposal we need to discuss three issues.

First, as foreshadowed in Sect. 1, let us consider the abstract idea of rational decision in light of one's beliefs and desires. Certainly, one way of making *irrational* decisions is to focus entirely on how much one desires certain outcomes and fail to factor in how likely one is to achieve them. Building on this insight, I take the fundamental guiding idea of rational decision making (henceforth FGI) to be that (a) one should choose between courses of action by considering how much one would like each of the possible outcomes to which they might lead, *and* (b) in this process one should weight how much one would like each outcome—one should factor it to a greater or lesser extent into one's decision whether to perform a certain action—according to the strength of one's belief that performing that action would lead to that outcome. So one does not simply throw the possible outcomes in together and consider which of them one would like best: one also tempers one's liking for each outcome according to the strength of one's belief that performing the action under consideration would lead to that outcome. This sort of idea goes right back to the origins of decision theory. It is a major theme of the final chapter of the *Port-Royal Logic*—for example:

in order to decide what we ought to do to obtain some good or avoid some harm, it is necessary to consider not only the good or harm in itself, but also the probability that it will or will not occur, and to view geometrically the

²⁷ 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.

proportion all these things have when taken together. (Arnauld & Nicole, 1996, 273–4)²⁸

Second, consider the distinction between *picking* and *choosing* as sub-types of the general act of *selection* [Ullmann-Margalit and Morgenbesser (henceforth UMM), 1977]. Choosing occurs when your reasons—your beliefs and desires—guide you to the item you select. For example, you choose to buy oats rather than frosted flakes because you believe they are cheaper and healthier and these considerations outweigh the fact that you prefer the taste of frosted flakes. Picking occurs when you have reason to select one of a group of items but no reason to select one particular member of the group rather than any of the others. For example, you have reason to buy a single box (that is all you can afford and all you can eat before they spoil) of a certain brand (your preferred brand) of oats but no reason to select any particular box from the shelf (that contains numerous boxes, all equally accessible, undamaged and with the same use-by date). In this case, you pick a box. The fact that we often find ourselves needing to pick is not a problem—for (as UMM make clear) we *can* pick. And while it is evident from the above explanation of what picking is that picking is in a clear sense *arbitrary*, we should be careful to strip this word of the negative connotations that it often carries, because (as UMM [768, 771] also make clear) picking is not *unreasonable*. The very fact that one has reason to pick *A* or *B* (rather than *C* or *D*) but no reason to pick *A* over *B* (or *B* over *A*) means that one is immune from regret if one picks *A* (or *B*).

Third, we need to consider utility. Specifically, I shall suppose that we can measure utility on (at least) an interval scale. That is, we have a numerical function representing an agent's desires for outcomes and it is (at least) unique up to positive linear transformation. This assumption is standard in the context of precise probability models—and it becomes no more problematic when we move to belief functions. Utilities could—for example—be identified with dollar values, or—better—derived by the method of von Neumann and Morgenstern (1953). The latter method makes use of the notion of *combinations* of events with given probabilities. For example, the 50–50 combination of *B* and *C* “is the prospect of seeing *B* occur with a probability of 50% and (if *B* does not occur) *C* with the (remaining) probability of 50%” [p.18].²⁹ This poses no problem in the context of a belief function representation of degrees of belief because precise probabilities are special cases of belief functions. The required assumption is that an agent's degree of belief that she will get *B* in the event of choosing an *x*–*y* combination of *B* and *C* is *x*—and her degree of belief that she will get *C* is *y*. In a precise probability framework, this means that the agent's probabilities for the outcomes of lotteries cannot be taken to be freely chosen or entirely subjective³⁰: they are taken to match the values that define the lottery. In a belief function framework, the required assumption is

²⁸ Note that the first three (comma-separated) parts of this quotation begin laying out the general idea of rational decision-making in light of one's beliefs and desires, while the fourth part not only finishes laying out this general idea, but also begins setting out the specific form this idea takes in a context in which beliefs are represented as probabilities—i.e. expected utility theory (see further n.31 below).

²⁹ These combinations are often called ‘lotteries’ in the literature. See e.g. Resnik (1987).

³⁰ Cf. von Neumann and Morgenstern (1953, 19).

essentially the same: that an agent's degrees of belief regarding whether she will ultimately obtain B or C in the event of choosing a certain combination of them are modelled by a Bayesian belief function whose values match the probabilities that define that combination. This is as unproblematic in the context of a belief function model of degrees of belief as it is in the context of a probabilistic model.

With the preliminaries dealt with, the question now arises how to implement FGI in the context of belief functions. In the classical context involving probabilities, and utilities of the sort just discussed, FGI is implemented by multiplying utilities by probabilities, and then adding the results.³¹ This is the familiar calculation of expected utility. In the belief function context, given utilities of the sort just discussed, the most straightforward way to (try to) implement FGI would be to go through an exactly analogous calculation, with the belief function in place of the probability function. So suppose we have mutually exclusive and jointly exhaustive states S_1, \dots, S_n and available acts A_1, \dots, A_q , where performing act A_i in state S_j leads to outcome $O_{i,j}$. We have utilities $U(O_{i,j})$ assigned to each outcome and we have a belief function Bel over the frame of discernment $\{S_1, \dots, S_n\}$. We then assign each act A_i an expected value $E(A_i)$ as follows³²:

$$E(A_i) = \sum_{j=1}^n Bel(\{S_j\}) \cdot U(O_{i,j}) \quad (5)$$

We then pick an act so as to maximise this value. In the special case where the belief function is Bayesian, this just *is* the familiar process of maximising expected utility.

This decision method seems not to have been considered in the literature on decision-making with belief functions. Technically it works, in the sense that setting $A_i \succcurlyeq A_r$ iff $E(A_i) \geq E(A_r)$ yields a complete preorder of the available acts (one that is invariant under positive linear transformations of the utilities) and a complete preorder is guaranteed to have at least one maximal element.³³ One then picks a maximal act. Nevertheless, the proposal cannot be accepted. Note for a start that this decision method disallows acts that from an intuitive point of view seem manifestly rational. Suppose that you are playing a cup game involving three cups, under one of which there is a ball. We can represent the frame of discernment as $\{C_1, C_2, C_3\}$, where C_i represents the possibility that the ball is under cup i . Suppose that your evidence warrants a basic mass assignment of 0.01 to $\{C_1\}$ and 0.99 to $\{C_2, C_3\}$. So you're nearly certain the ball is under cup 2 or 3 but have no further information regarding which of these cups it is specifically. Now suppose that you have available four acts, or bets (Table 1). Bet 1 pays \$1 on C_1 and nothing otherwise; bet 2 pays \$1 on each outcome; bet 3 pays \$1 on C_1 and \$10 on each of C_2 and C_3 ; and bet 4 pays 99¢ on C_1 and \$100 on each of C_2 and C_3 . By (5), bets 1, 2 and 3 have the same

³¹ I take it that this mode of combination is what is meant by 'geometrically' in the quotation from the *Port-Royal Logic* above—hence my comment in n.28 that the quotation covers both the abstract idea of rational decision in light of beliefs and desires and its particular implementation in a probabilistic context.

³² I think 'expected value' is the most useful term to use here but note that this value is not in general (as in the precise probability case) the expectation of a random variable.

³³ A binary relation \succcurlyeq is a preorder iff it is reflexive and transitive. It is complete iff for all x and y , $x \succcurlyeq y$ or $y \succcurlyeq x$. A maximal element x is one such that for every y , $x \succcurlyeq y$.

Table 1 Cup game mark I

belief mass	ball under	bet 1	bet 2	bet 3	bet 4
0.01	cup 1	\$1	\$1	\$1	99¢
0.99	cup 2	\$0	\$1	\$10	\$100
	cup 3	\$0	\$1	\$10	\$100

expected value, and it is greater than that of bet 4.³⁴ Intuitively however it would be acceptable—if not mandatory—to prefer bet 2 to bet 1, bet 3 to bet 2, and bet 4 to bet 1.

Moving from intuitions about a case to a more theoretical level, the problem is that the decision method simply ignores some of your beliefs. It attends only to degrees of belief in singleton subsets of the frame of discernment. The strong evidence, and corresponding strong degree of belief you have for the (non-singleton) subset $\{C_2, C_3\}$, never comes into play. The decision method is ranking bets entirely according to their utilities on outcome C_1 , because that is the only outcome for which you have a non-zero degree of belief. It is thereby ignoring the fact that you have strong evidence and a corresponding high degree of belief that the outcome will in fact be C_2 or C_3 .

Indeed, I want to argue further that this decision method does not accord with FGI. On the one hand, FGI is, as its name suggests, a guiding idea. It is not supposed to equate to a specific mathematical procedure, such as maximising expected utility. Rather, maximising expected utility is supposed to be a way of implementing FGI in a context in which beliefs are formally modelled using probability functions (and desires using utility functions)—a way of making an abstract, guiding idea concrete in that setting. On the other hand, FGI is not supposed to be an empty metaphor. Not just *any* decision method that takes two inputs—one representing beliefs and one representing desires—should automatically count as an implementation of FGI. And I shall now argue that the method currently under consideration does not—even though it does indeed have two inputs: a belief function, modelling belief; and a utility function, modelling desire. The problem with the method is that it does not appropriately weight desires by degrees of belief. To see this, consider a series of variations on the cup game (Table 2). Bet 1 pays out $\$x$ on cup 1 and nothing otherwise. Bet 2 pays a prize $\$y$ —bigger than $\$x$ —if the ball is under cup 2 or 3, and a consolation prize $\$z$ —slightly smaller than $\$x$ —if the ball is under cup 1. Suppose that your evidence warrants a basic mass assignment of some degree of belief k to $\{C_1\}$ with $1 - k$ going to $\{C_2, C_3\}$. Now suppose we gradually lower x (but keep it greater than 0), raise y , lower z to keep it just under but ever closer to x , and lower k (but keep it non-zero). My claim is that a decision method is not according with FGI if, in this process, it never stops deeming bet 1 to be superior to bet 2. If the method just fixates on bet 1 and deems it the best act for you to choose, no matter how low x and k go nor how high y goes, then it is not appropriately factoring in just

³⁴ Sometimes, for ease of presentation, I assume—inessentially—that utilities match monetary values.

Table 2 Cup game mark II

belief mass	ball under	bet 1	bet 2
k ($k > 0, k \rightarrow 0$)	cup 1	x ($x > 0, x \rightarrow 0$)	z ($z = x - \Delta, \Delta > 0, \Delta \rightarrow 0$)
$1 - k$ {	cup 2	0	y ($y > x, y \rightarrow \infty$)
	cup 3	0	y

how much you like the prize y —how much more you like it than x —and how unlikely you think the ball is to be under cup 1. It is not appropriately weighting your desires by your degrees of belief. And the decision method currently under consideration fails this test. In all these cases it assigns bet 1 expected value $\$kx$ and bet 2 a strictly lower expected value $\$kz$.

4 Deciding with belief functions

As mentioned in Sect. 2, for probability functions, assignments to singleton subsets (or equivalently elements) of the frame of discernment determine the entire function, but this is not the case for belief functions. So really it is not surprising that copying the probabilistic expected utility calculation—where degrees of belief in singleton subsets of the frame are multiplied by utilities of corresponding outcomes—yields a poor result. The analogues of the singleton sets in the belief function case are the focal sets: mass assignments to these *do* determine the entire belief function. But if we try to implement FGI by multiplying basic mass assignments by utilities, we face a correlative problem. The utility function assigns values to outcomes determined by acts together with particular states in the frame of discernment—not whole sets of states. The problem, then, is that FGI requires us to weight desires by beliefs—but in general the two do not mesh. Generally speaking, beliefs are clump-like: the evidence can warrant attaching positive belief only to non-singleton subsets of the frame of discernment. Desires, on the other hand, are finely individuated: utilities attach to particular outcomes, determined by actions together with individual elements (or equivalently singleton subsets) of the frame of discernment.³⁵ To see our

³⁵ I am assuming a framework in which utilities are assigned to particular outcomes. Such a framework is found in Savage (1972). There is a different kind of framework, found in Jeffrey (1983), in which utilities are assigned to propositions. I could argue for the preferability of the Savage-style approach in general, but that is beyond the scope of this paper. In the present context in particular, however, it might seem that the Jeffrey-style approach would avoid the problem outlined above, because it involves assigning utilities to all propositions, including the ‘clumpy’ (nonspecific) ones corresponding to the focal sets of the belief function. When we look closely, however, there is no advantage here. Jeffrey’s desirabilities of nonspecific propositions are really (as he notes himself on pp. 86–7) *expected* desirabilities, determined by the desirabilities and probabilities of the more specific ways in which they can come true. But the present problem is precisely that we can have positive degrees of belief in certain subsets of the sample space *without* having non-zero degrees of belief in any of their subsets. For Jeffrey, if the probabilities of X and Y are zero, then the desirability of $X \vee Y$ will also have to be zero (see his axiom (5–2) on p.80).

way to a general solution to this problem, let's start by considering two special cases where things work straightforwardly.

The first case is where the belief function is Bayesian. In this case, focal sets are singletons: so beliefs are not clumpy—they are as finely individuated as utilities—and we can weight desires by beliefs by multiplying utilities by degrees of belief in the classical way. The second case is where the utility function assigns the same value to all outcomes of an act determined by states within the same focal set. For example, suppose that you are considering a die roll—the frame of discernment is $\{1, 2, 3, 4, 5, 6\}$ —and your evidence (a report, with some pages missing, of the methods used to produce trick dice at a games factory) leads you to assign mass 0.3 to $\{2, 4, 6\}$ and 0.7 to $\{1, 3, 5\}$. But suppose that the bet you are considering pays out the same amount (say \$1) for all even rolls and the same amount (say \$2) for all odd rolls. In that case, you can proceed as if 'even' is a state that, together with the act of accepting the bet, leads to an outcome—getting \$1—to which you assign a particular utility; and similarly for 'odd'. The point is that as far as you care—as far as your utility function is concerned—there is no difference between 2, 4 and 6. So you can proceed as if the utility of \$1 attaches to 'even' itself (in concert with the act of accepting the bet)—i.e. to the event about which you have a positive degree of belief—not to the individual states (2, 4 and 6) about which you have no evidence and hence no positive degree of belief. And then you can weight desires by beliefs by multiplying these utilities (of 'even' and 'odd') by your degrees of belief in those (non-singleton) events (and summing the results)—i.e. performing something like an expected utility calculation, but at the level of focal sets rather than elements (or singleton subsets) of the sample space. In the current example, your expected value for the bet will be $0.3 \times 1 + 0.7 \times 2 = 1.7$, that is the sum of your degree of belief in 'even' multiplied by its utility, and your degree of belief in 'odd' multiplied by its utility.

So things work out nicely when your utilities and your degrees of belief have the same granularity: either your beliefs distinguish as finely as your finely cut utilities (as in the first case above); or your utilities do not differentiate any more finely than your clumpy degrees of belief (as in the second case above). So what to do in the general case? For example, suppose your evidence about the die is as above, but the bet pays out differently on 2, 4 and 6 (and on 1, 3 and 5). How to decide whether to take the bet (at a certain price)? You have beliefs (modelled by a belief function) and desires (modelled by a utility function) and FGI tells you to decide by weighting your desires by your beliefs. However, you cannot do this because your beliefs and desires do not mesh. So first, you have to get them to mesh: you have to connect them up to one another. I propose that you do this via a proportioning of the focal sets of your belief function.³⁶ With a proportioning in hand, there are two ways to proceed: divide up your beliefs; or pool your utilities. These two approaches should, in my view, be regarded as two ways of spelling out the same process—meshing your beliefs and desires so as to be able to weight the latter by the former—rather than two fundamentally different processes. As we shall see, for each way of doing one, there

³⁶ The idea of a proportioning was introduced in Sect. 2.

is a way of doing the other that leads to the same assignments of expected values to actions.

The first approach is to use the proportioning to divide basic masses attached to non-singleton sets amongst their singleton subsets. As discussed in Sect. 2, this generates a Bayesian belief function—a probability function. You then assign an expected value to each act A_i by performing a standard expected utility calculation (using this probability function).

The second approach is to use the proportioning to pool your utilities for outcomes generated by an act together with states in the same focal set, to generate a utility value for the entire focal set (together with the act in question). Suppose (as previously) the frame of discernment is $\{S_1, \dots, S_n\}$, the focal sets of your basic mass assignment m are S_1, \dots, S_k , and the utility assigned to the outcome O_{ij} of act A_i in state S_j is $U(O_{ij})$. Given a proportioning Pr_1, \dots, Pr_k of the focal sets S_1, \dots, S_k , you attach a utility $U_{S_p}^i$ to each pair of an act A_i and a focal set S_p as follows:

$$U_{S_p}^i = \sum_{S_j \in S_p} Pr_p(S_j) \cdot U(O_{ij}) \tag{6}$$

The general idea here is (as mentioned) to obtain a utility for the focal set (together with the act under consideration) by pooling the utilities of outcomes determined by states in the focal set (together with the act under consideration)—and the particular method of pooling is that each state contributes its own proportion (of the focal set in question, as given by the proportioning under consideration) of the utility of the outcome to which it (together with the act under consideration) would lead. Less precisely, but more simply: each state in the focal set contributes *its* proportion of *its* utility; adding these together gives a utility for the focal set as a whole. With utilities for focal sets in hand, you assign an expected value to each act A_i using the basic idea already illustrated above (in the case of the bet that pays out the same amount for all even rolls and the same amount for all odd rolls):

$$E(A_i) = \sum_{p=1}^k m(S_p) \cdot U_{S_p}^i \tag{7}$$

The two approaches just presented yield the same expected values for acts (provided the same proportioning is used). Consider the first approach. As explained in Sect. 2, a division of basic masses among singleton subsets of the focal sets generates the following probabilities for singletons³⁷:

$$P(\{S_j\}) = \sum_{p=1}^k Pr_p(S_j) \cdot m(S_p) \tag{8}$$

We then calculate expected values for acts by taking the standard classical equation for the expected utility of A_i :

³⁷ In Sect. 2 this was presented in two steps: first we defined $m_{p,j} = Pr_p(S_j) \cdot m(S_p)$; second (using the notation m' , which in (8) has been replaced with P) we defined $m'(\{S_j\}) = \sum_{p=1}^k m_{p,j}$.

$$E(A_i) = \sum_{j=1}^n P(\{S_j\}) \cdot U(O_{i,j}) \quad (9)$$

and plugging in the probabilities from (8):

$$E(A_i) = \sum_{j=1}^n \left(\sum_{p=1}^k Pr_p(S_j) \cdot m(\mathbf{S}_p) \right) U(O_{i,j}) \quad (10)$$

Now consider the second approach. Plugging (6) into (7) we see that it assigns the following expected values:

$$E(A_i) = \sum_{p=1}^k \left(m(\mathbf{S}_p) \sum_{S_j \in \mathbf{S}_p} Pr_p(S_j) \cdot U(O_{i,j}) \right) \quad (11)$$

To facilitate comparison between this and (10), note that it wouldn't make any difference to (6) if instead of summing over all S_j in \mathbf{S}_p , we summed over all S_j in the frame of discernment, provided we define $Pr_p(S_j) = 0$ when $S_j \notin \mathbf{S}_p$ (cf. n.22). This gives us the following version of (6)³⁸:

$$U_{\mathbf{S}_p}^i = \sum_{j=1}^n Pr_p(S_j) \cdot U(O_{i,j}) \quad (12)$$

If we plug (12) (instead of (6)) into (7), we get the following (instead of (11)):

$$E(A_i) = \sum_{p=1}^k \left(m(\mathbf{S}_p) \sum_{j=1}^n Pr_p(S_j) \cdot U(O_{i,j}) \right) \quad (13)$$

We can now easily compare (13) and (10) and see that they are equivalent: we get from one to the other by expanding, commuting and factorising.³⁹

My proposal, then, is this. To make a decision in light of your beliefs—represented by a belief function—and your desires—represented by a utility function—you first pick a way of getting them to mesh—a proportioning of the focal sets of the belief function—and then maximise expected value. There are two equivalent ways of calculating the expected values of acts: use the proportioning to generate a probability function from your belief function, and then maximise expected utility in the classical way; or use the proportioning to pool your utilities (for outcomes determined by the act under consideration) into utilities for focal sets, and then do something analogous to the classical calculation of expected utility, except that it involves a basic mass assignment (not a probability function) and utilities determined

³⁸ It also means that in (8), we can drop the qualification that we sum the $Pr_p(S_j) \cdot m(\mathbf{S}_p)$ that are indeed defined (see n.23): if we set $Pr_p(S_j) = 0$ when $S_j \notin \mathbf{S}_p$ then they are all defined.

³⁹ Evidently one could also specify intermediate or hybrid ways of proceeding that yield the same results. Instead of dividing the belief masses assigned to focal sets among their singleton subsets, one could divide them among (not necessarily singleton) subsets and then pool utilities to determine utility values for these subsets.

by (the act under consideration and) sets of states—i.e. focal sets—rather than individual states in the frame of discernment.

Two things are worth noting about the first step—picking a proportioning. First, given FGI, it cannot be a matter of rational decision. Rational decision involves weighting your desires by your beliefs. The proportioning is required in order for you to do this weighting: it is a necessary precursor. It has to come *first* and *then* you can weight things up. So there can be no question of which proportioning you should rationally choose: you have to *pick* one. Second, the proportioning should not be seen as a third element in decision-making, in addition to belief and desire. It is a way of linking the two elements in decision-making—belief and desire—so that the former can be used to weight the latter, in accordance with FGI. Consider the process of calculating expected utility in a classical context where beliefs are represented by probabilities. We multiply probabilities by utilities and then add the results. In my view, we should not ask what multiplication and addition *mean* in this context or what psychological processes they model (and indeed these questions are not asked, as far as I am aware). Rather, I view the situation as follows. FGI tells us to weight our desires by our degrees of belief. We should be more (less) attracted to an act the more (less) we like the outcomes to which it might lead—but this is not all: we should factor in our liking for these outcomes more or less according to the strength of our belief that performing the act will lead to them. Now the point here is not to get fixated on the idea of ‘weighting’ and then think that this *means* multiplying and adding. Rather, the key point is that when we stand back and observe the results, it seems clear that calculating expected utility in the classical way does cohere with FGI: it is indeed a way of rating actions more (less) highly if we believe more (less) strongly that they will lead to outcomes that we desire more (less). Now in the situation where beliefs are modelled by belief functions, we can—mathematically speaking—perform an analogous calculation, in which we multiply beliefs by desires and add these products. However when we stand back and observe the results it seems clear that they do not (in this new setting) cohere with FGI. That was the point made at the end of Sect. 3. So we need to use different mathematical machinery. My proposal is that we first use a proportioning—and *then* multiply and add. The point that I am making now is that the proportioning should be seen alongside the multiplication and the addition, not alongside the belief and desire. It is not a third ingredient in rational decision-making but part of the mathematical machinery (the other parts being multiplication and addition) needed to bring the two ingredients (belief and desire) together in a way that coheres with FGI.

To make sure that the proposal is clear, let us run through it again, but this time in a less abstract way: focusing specifically on the first method of proceeding, where you generate a probability function from your belief function (and then maximise expected utility). In this setting, the essential line of thought is as follows. You don’t have enough evidence to warrant a probabilistic belief state. That is why your beliefs are represented by a (non-Bayesian) belief function. But to implement FGI by maximising expected utility (in the classical way), you need probabilistic degrees of belief: we saw in Sect. 3 what goes wrong if you try to do it with the belief function

directly.⁴⁰ If you had more evidence—enough to warrant a Bayesian belief function—you could implement FGI by maximising expected utility. So you need to proceed as if you have further information or evidence: enough to determine a probabilistic belief state. However, there is no particular piece of further evidence—no particular probabilistic belief state—you should act as if you have. After all, the whole point is that your evidence does not warrant a probabilistic belief state: it is modelled by a belief function that is *not* a probability function and with which *many* probability functions are compatible. So you have to *pick* a compatible probability function—and then maximise expected utility. In accordance with FGI, you are using your beliefs to weight your desires. You are not throwing them away and acting as if you had completely different beliefs. You are augmenting them with further detail—you are picking a *compatible* probability function. You use the beliefs you have by pretending they are more detailed than they really are—so that they mesh with your desires in a way that enables an expected utility calculation. Recall from Sect. 2 the idea of seeing the lines connecting the black regions to the central (non-)Venn diagram as scale arms, and the black regions as pans, over which you distribute a unit of belief mass. Relative to each possible act that you might perform, an outcome appears on each pan that represents a *singleton* subset of the frame of discernment: if you perform that act, and the state in that singleton is the one that turns out to be actual, then that is the outcome you will obtain. You have utilities for each of these outcomes. To choose an act, FGI tells you not just to consider these utilities—how much you would like each outcome—but to weight them according to your degrees of belief. Your problem is that you do not necessarily have a belief mass associated with each singleton subset: sometimes you have a chunk of belief mass sitting on a pan representing a larger set, with nothing on any of the pans representing its singleton subsets. My proposal is that you proceed by breaking this chunk into parts and distributing them to the pans representing the singleton subsets—i.e. in effect you reassemble the exploded Venn diagram. *How* you do this cannot be rationally mandated by FGI because it is a *precursor* to the kind of weighting that FGI demands. You can do it however you like: you *pick* a way of dividing the mass amongst the singletons. But note that you are still using your beliefs to weight your desires. All you may do is break the belief mass into chunks and distribute them amongst singleton subsets. You may not add any mass; you may not discard any; and you may not distribute any to non-subsets. Within these limits, there is no particular mandated way for you to proceed: but proceeding in any of these ways is still using your beliefs to weight your desires. While at the formal level you derive a probability function from the belief function, it is crucial that what is going on at the conceptual level is that you are *using* your non-probabilistic degrees of belief (to weight your desires, in accordance with FGI)—not *changing* them. You can be described as bringing your degrees of belief to bear by *acting as if* they are probabilistic and you are maximising expected utility—but the beliefs themselves remain non-probabilistic. Your degrees of belief are (still) represented by a (non-Bayesian) belief function

⁴⁰ In Sect. 5 we'll see why a second possible way of using the belief function directly also fails.

because that is what the evidence warrants (and in general the fact that you need to decide and then act provides no new evidence).⁴¹

5 Comparison with other views

In this section, I consider alternatives to the proposal of the previous section. These views are drawn not only from the literature on belief functions, but also from the literatures on decisions under ambiguity, and on imprecise probabilities.⁴² The reason for this wide scope is that—as discussed in Sect. 2—a belief function determines a set of compatible probability functions, so methods for deciding in light of a set of probability functions are also relevant.

First consider the Transferable Belief Model (TBM): a two-level model comprising “a *credal level* where beliefs are entertained and a *pignistic level* where beliefs are used to make decisions” (Smets & Kennes, 1994, 192). The credal level is modelled by a belief function. The pignistic level is modelled by a probability function that is obtained from the belief function by a *pignistic transformation*, when a decision needs to be made. Specifically, this probability function is what we may call the Laplacean probability: in the terms of this paper, it is obtained from the proportioning which gives an *equal* share of each focal set to each of the set’s members.⁴³ The agent then makes a decision by maximising expected utility relative to this probability function. The proposal presented in Sect. 4 is that one picks a proportioning, which can then be used to generate a probability function or to pool utilities, with the resulting expected values of acts being the same either way—and then maximises expected value. However, here (and in other places below), it will facilitate comparison if we present the view in a more simplistic way: pick a compatible probability function and then maximise expected utility. The TBM differs in that it requires us always to use the Laplacean probability: it replaces the picking step in my approach with a mandated choice of a particular probability function. Obviously decisions recommended by the TBM will be allowable on my view—but the converse need not hold. That is, the TBM deems unacceptable some decisions allowed on my view: certain decisions made by maximising expected utility after picking a probability other than the Laplacean one. But why think they are unacceptable? Why think that when we generate a probability function from our belief function, for purposes of making a decision, we must choose the Laplacean probability, rather than picking any compatible probability? I have already argued that this does not follow from FGI. But if a mandated choice of probability function

⁴¹ My view is thus different from that of Moss (2015), who—in the context of imprecise probabilities (and incommensurable values)—writes that “any rational agent with this sort of imprecise mental state identifies with some precise mental state for purposes of action” (p.673; cf. also p. 675). In my view, the agent *acts as if* she has probabilistic beliefs but she does not *identify* with any precise probabilistic belief state.

⁴² For overviews, see Denœux (2019), Etner et al. (2012), and Troffaes (2007) and Huntley et al. (2014).

⁴³ See e.g. Williams (1982, 342), Dubois and Prade (1986, 214) and Smets and Kennes (1994, 201–2). Cf. Laplace (1951, 6): “We know that of three or a greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. ...The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence”.

doesn't follow from FGI—from considerations of pure rationality—then the requirement must be a pragmatic one. However by their nature, pragmatic requirements are local and context-dependant: they are tuned to practical considerations, and these vary from context to context. I can certainly accept that when considering decisions made in a specific kind of context, someone might want to bolt on to the view of this paper the further requirement that one choose a certain probability function (or more generally, a certain proportioning)—rather than picking any compatible one. Indeed I consider it a strength of my proposal that while in general the first step—selecting a proportioning—is an act of picking, it *can* in certain contexts be replaced by a mandated choice of a particular proportioning (or compatible probability function), for local pragmatic reasons. For example, decisions made in certain public organisations, or decisions made in similar contexts across multiple sites within an organisation, might need to be systematic, predictable and repeatable. Or in another kind of context, it could be that certain mathematical properties are important. For example, where m_1 and m_2 are basic mass assignments and F is a transformation of basic mass assignments into probability functions, F is said to be *linear* iff for any $\alpha \in [0, 1]$:

$$F(\alpha.m_1 + (1 - \alpha).m_2) = \alpha.F(m_1) + (1 - \alpha).F(m_2)$$

Smets (2005) shows that the only transformation F that is linear is the Laplacean transformation (and that the Laplacean probability it yields corresponds to the Shapley (1953) value in cooperative game theory)—and I can certainly accept that this property could be important in certain contexts. However, at a more abstract level—at the level of giving a general decision theory for belief functions that respects requirements of pure rationality and does not impose any further restrictions, which might be relevant in particular contexts but cannot be justified in general—there is no good reason for replacing the picking step of my decision procedure with a mandated choice of one particular probability function.⁴⁴

Second consider a view that, like the TBM, mandates a particular choice of probability function—but a different one: not the Laplacean probability, but the *plausibility probability* defined by assigning to each singleton subset its plausibility PI (see n.15) divided by a normalisation constant equal to the sum of the plausibility values of all singleton subsets of the frame of discernment (Cobb & Shenoy, 2003, 2006). The problem with this proposal, from my point of view, is that the plausibility probability need not be compatible with the belief function. For example, suppose that in a cup game of the sort considered above your evidence warrants a basic mass assignment of 0.5 to $\{C_1\}$ and 0.5 to $\{C_2, C_3\}$. In this case, $PI(\{C_1\}) = PI(\{C_2\}) = PI(\{C_3\}) = 0.5$ and the total sum of plausibility values of singleton subsets of the frame of discernment is 1.5. Hence, the plausibility probability assigns $\frac{1}{3}$ to each of the possibilities C_1 , C_2 and C_3 . This is not compatible with the original belief function, which demands that C_1 get probability 0.5. If you decide by

⁴⁴ In particular, while Smets and Kennes (1994, 202–3) derive their pignistic probability from certain axioms, these axioms are not (and, to be clear, are not claimed to be) all compelling as general rationality requirements.

maximising expected utility relative to the plausibility probability, you will be underweighting possibility C_1 and departing from FGI.

Similar comments apply to any other view that mandates a choice of a single probability function. If the probability function is compatible with the belief function, then the proposal is guilty of overreach: of mandating a choice where FGI leaves the agent in a picking situation. If the probability function is not compatible with the belief function, then the proposal violates FGI.

The views to consider next can all be presented in terms of one or both of the lower and upper expectations of an act. Instead of first picking a probability and then maximising expected utility, for each act A_i we consider *all* the expected utility values one could derive for A_i by picking a compatible probability function and then calculating A_i 's expected utility (relative to that probability function). The infimum of these values is the lower expectation of the act A_i , denoted $\underline{E}(A_i)$, and the supremum of these values is the upper expectation of A_i , denoted $\overline{E}(A_i)$. These lower and upper expectations can equivalently be defined without reference to compatible probability functions:

$$\underline{E}(A_i) = \sum_{p=1}^k m(\mathbf{S}_p) \cdot \min\{U(O_{i,p_1}), \dots, U(O_{i,p_p})\} \tag{14}$$

$$\overline{E}(A_i) = \sum_{p=1}^k m(\mathbf{S}_p) \cdot \max\{U(O_{i,p_1}), \dots, U(O_{i,p_p})\} \tag{15}$$

Note the similarity and the difference between (14) and (15), and the second way of presenting my approach in Sect. 4, where we use the proportioning of the focal sets to pool utilities into a utility value for the entire focal set and then calculate an expected value using (7). The similarity is that (7), and (14) and (15), all fit the format:

$$E(A_i) = \sum_{p=1}^k m(\mathbf{S}_p) \cdot U_{\mathbf{S}_p} \tag{16}$$

where $U_{\mathbf{S}_p}$ is a function of $U(O_{i,p_1}), \dots, U(O_{i,p_p})$. In (14) $U_{\mathbf{S}_p}$ is min. In (15) $U_{\mathbf{S}_p}$ is max. In (7) $U_{\mathbf{S}_p}$ is $U_{\mathbf{S}_p}^i$ as defined in (6). The difference is that $U_{\mathbf{S}_p}^i$, unlike min and max, does not operate directly on the utilities $U(O_{i,p_1}), \dots, U(O_{i,p_p})$: it also takes as input the proportions of \mathbf{S}_p assigned to each $S_j \in \mathbf{S}_p$.⁴⁵

With the lower and upper expectations in hand, we can specify a number of different decision methods. The first is (generalised) maximin: pick an act that maximises \underline{E} .⁴⁶ My criticism of this view runs along similar lines to my criticism of the view considered in Sect. 3 and reaches the same conclusion: the view does not cohere with FGI. Consider another series of variations on the cup game (Table 3).

⁴⁵ See further the discussion of the OWA view below.

⁴⁶ See Gärdenfors and Sahlin (1982), Berger (1985, Sect. 4.7.6) and Gilboa and Schmeidler (1989). Cf. the maximin method for decisions under ignorance (Wald, 1945).

Table 3 Cup game mark III

belief mass	ball under	bet 1	bet 2	bet 3
k ($k > 0, k \rightarrow 0$)	cup 1	x ($x > 0, x \rightarrow 0$)	0	0
$1 - k$ {	cup 2	0	y ($y > x, y \rightarrow \infty$)	0
	cup 3	0	0	y

Suppose we gradually lower x (but keep it greater than 0), raise y , and lower k (but keep it non-zero). My claim is that a decision method is not according with FGI unless, in this process, it at some point deems bet 2 or 3 acceptable. If it just fixates on bet 1 and deems it the best act for you to choose, no matter how low x and k go nor how high y goes (imagine for example that k is 0.0001, bet 1 pays out one cent, and bets 2 and 3 pay out one million dollars), then it is not appropriately factoring in just how much you like the prize y —how much more you like it than x —and how incredibly unlikely you think the ball is to be under cup 1. It is not appropriately weighting your desires—your increasingly strong desire for y and increasingly weak desire for x —by your degrees of belief—your increasingly strong degree of belief that the ball is not under cup 1 and is under cup 2 or cup 3. But maximin does not do this: it *always* tells you to pick bet 1 over bets 2 and 3. This is because the former always has non-zero worst-case expected value (kx), whereas the latter both have zero worst-case expected value (when all the $1 - k$ belief originally accorded to $\{C_2, C_3\}$ is applied to $\{C_2\}$, bet 3 has expected value 0; and when it is all applied to $\{C_3\}$, bet 2 has expected value 0).⁴⁷

Another option is (generalised) maximax: pick an act that maximises \bar{E} . The problem with this approach is that it too fails to cohere with FGI. Again consider a series of variations on the cup game (Table 4). There are n cups. You have degree of belief 0.5 that the ball is under cup 1, and 0.5 that it is under one of the remaining cups. Bet 1 pays out on cup 1 (only), and bet 2 pays out slightly more on cup n (only). Now suppose we gradually increase the number of cups and reduce the difference between bet 2's payout and bet 1's payout. My claim is that a decision method is not according with FGI unless, in this process, it at some point deems bet 1 acceptable. (Imagine for example that n is 1000, bet 1 pays out one million dollars, and bet 2 pays out one cent more.) For note that as the number of cups increases, your 0.5 degree of belief that the ball is not under cup 1 becomes increasingly indefinite, in the sense that it leaves open more and more possibilities (more and more options as to where the ball is—more and more cups it could be under). Yet maximax completely ignores this: no matter how many cups there are, it proceeds as if the 0.5 rests on cup n (the cup on which bet 2 pays out) and demands that you prefer bet 2 to bet 1. My view *allows* you to assign the 0.5 to cup n and hence prefer bet 2, but it does not demand this: it also allows you to distribute the 0.5 evenly over

⁴⁷ See below for further discussion.

Table 4 Cup game mark IV

belief mass	ball under	bet 1	bet 2
0.5	cup 1	$x \quad (x > 0)$	0 on all cups $< n$ $y \quad (y = x + \Delta, \Delta > 0, \Delta \rightarrow 0)$
0.5	cup 2	0 on all cups > 1	
	\vdots cup $n \quad (n \geq 2, n \rightarrow \infty)$		

Table 5 Cup game mark V

belief mass	ball under	bet 1	bet 2
0.5	cup 1	\$1 on all cups	50¢ on all cups $< n$ \$2.54
0.5	cup 2		
	\vdots cup $n \quad (n \geq 2, n \rightarrow \infty)$		

all of cups 2 through n and hence prefer bet 1 when there are three (or more) cups (assuming bet 2’s payoff is less than twice that of bet 1).

A third option is (generalised) Hurwicz or α -maximin: decide by maximising $\alpha \underline{E} + (1 - \alpha) \overline{E}$ for some $\alpha \in [0, 1]$.⁴⁸ I have already argued against this in the cases where α is 1 (which gives us maximin) or 0 (which gives us maximax)—but even setting α at 0.5 does not guarantee satisfactory results. Consider another series of variations on the cup game (Table 5). α -maximin demands that you prefer bet 2 (where \underline{E} is 50¢ and \overline{E} is \$1.52 and hence $\alpha \underline{E} + (1 - \alpha) \overline{E}$ is \$1.01 when $\alpha = 0.5$) over bet 1 (where $\alpha \underline{E} + (1 - \alpha) \overline{E}$ is \$1). This is the case *no matter how many cups there are*. But as already discussed, the more cups there are, the more indefinite your 0.5 degree of belief that the ball is not under cup 1 becomes. The present view just fixates on your desire for \$2.54 and does not allow any room for your increasingly indefinite degrees of belief to alter your decision. (Imagine for example that there are a billion cups—and note that bet 2 pays out \$2.54 on only a single one of them.) The view does not involve appropriately weighting your desire by your degree of belief: it fails to implement FGI. (My view allows you to prefer bet 2 but unlike the current view does not require you to: for example, if there are four cups, and you divide your 0.5 belief that the ball is not under cup 1 evenly over the three open possibilities, then my view will recommend bet 1.)

The next class of views to consider comprises those that assign an expected value to each act in accordance with (16) where U_{S_p} in an ordered weighted averaging (OWA) operator, i.e. a function

⁴⁸ See Jaffray (1988) and Strat (1990). Cf. the Hurwicz (1951) method for decisions under ignorance.

$$F\langle U(O_{i,p_1}), \dots, U(O_{i,p_l}) \rangle = \sum_{q=1}^{l_p} w_q U_{(q)} \quad (17)$$

where w_1, \dots, w_{l_p} are non-negative weights that sum to 1 and $U_{(q)}$ is the q th largest of the utilities $U(O_{i,p_1}), \dots, U(O_{i,p_l})$.⁴⁹ That is, relative to an act A_i one assigns a utility to the focal set S_p by *ordering*, from largest to smallest, the utilities determined by act A_i together with *elements* of the focal set, applying a weight to each one, and then summing the results. Depending on the choice of weights, this framework can yield decision recommendations equivalent to those of views already discussed:

Weights	View
$\frac{1}{l_p}, \dots, \frac{1}{l_p}$	TBM
$0, \dots, 0, 1$	Maximin
$1, 0, \dots, 0$	Maximax
$1 - \alpha, 0, \dots, 0, \alpha$	α -maximin

From my point of view, then, some instances of the OWA view constitute unacceptable decision theories, while some yield acceptable (but not mandatory) recommendations. From a conceptual point of view, the problem (as I see it) with the OWA view is that if we first order the utilities *by size*, and then assign each a proportion or weight, we can end up in effect weighting desires by desires—not by beliefs, as FGI requires. This happens for example when we set weights so as to end up with recommendations equivalent to maximin, or maximax, which (respectively) allow outcomes you *desire* least, or most, to hold undue sway over your decision-making. Recall for example the version of the cup game in Table 3. Your evidence points you strongly towards cup 2 *or* 3. Yet maximin in effect has you proceed as if the ball will not be under cup 2 when you are considering bet 2, and as if it will not be under cup 3 when you are considering bet 3. Your dislike of getting nothing plays too much of a role here, and ends up washing out your strong belief that the ball will be under cup 2 or 3. The view of this paper, by contrast, is blind to any utilities associated with elements of the focal set, at the point at which it assigns each element a proportion: it only later multiplies the element's proportion by a utility. It is therefore not possible, on my view, to gerrymander the proportions so that the least-desired (or most-desired) element of the focal set invariably gets the greatest proportion.

The next view to consider is (generalised) minimax regret.⁵⁰ For each act A_i and each state S_j in the frame of discernment, we define the regret value R_i^j of that act relative to that state as the difference between the utility of the outcome determined by that act in that state, and the utility of the best outcome you could have got in that state by performing one of the available acts. Where \mathfrak{A} is the set of all available acts:

⁴⁹ See Yager (1992).

⁵⁰ See Yager (2004).

Table 6 Regret values for cup game mark III

belief mass	ball under	bet 1	bet 2	bet 3
k	cup 1	0	x	x
$1 - k$	cup 2	y	0	y
	cup 3	y	y	0

$$R_i^j = \max_{A_r \in \mathfrak{A}} U(O_{r,j}) - U(O_{i,j}) \tag{18}$$

We now define the regret value of act A_i relative to a set S of states by taking the maximum of its regret values relative to each of the states in the set:

$$R_i^S = \max_{S_j \in S} R_i^j \tag{19}$$

We can now assign to each act A_i an expected regret value $R(A_i)$:

$$R(A_i) = \sum_{p=1}^k m(S_p) \cdot R_i^{S_p} \tag{20}$$

The idea is then to pick an act so as to minimise this value. My criticism of this view runs along similar lines to my criticism of maximin. Recall the cup game from Table 3. My claim was that a decision method is not according with FGI unless it at some point deems bet 2 or 3 acceptable. Minimax regret fails this test. Table 6 shows the regret values for each act, obtained by subtracting each payoff in Table 3 from the highest payoff in its row. The regret value for each act relative to the set $\{C_2, C_3\}$ is its maximum regret across rows 2 and 3: so y for each bet. Thus bet 1’s expected regret is $k \cdot 0 + (1 - k)y = (1 - k)y$ and bet 2 and bet 3 each have expected regret $k \cdot x + (1 - k)y$. So minimax regret always demands that one prefer bet 1, because bet 1’s expected regret is always $k \cdot x$ less than the expected regrets of bets 2 and 3.

All the views considered so far follow this pattern: assign a *number* to each act and then maximise (or in the last case minimise) this number. Now we consider three views that fit a different pattern. First, we define a relation between acts. Then we define the choice set—i.e. the set of choiceworthy or acceptable acts (from which the decision-maker is to pick one, if the choice set contains more than one act)—as the set of all nondominated acts (with respect to this relation): i.e. act A_i is choiceworthy iff there is no available act A_r such that A_r stands in the relation to A_i but not vice versa. The first two views that fit this pattern define their relations as follows (using the notions of lower and upper expectations)⁵¹:

⁵¹ For the first view see Satia and Lave (1973), Kyburg (1983), Ramoni and Sebastiani (2001) and Zaffalon et al. (2003); for the second view see Destercke (2010).

$$A_i \supseteq A_r \Leftrightarrow \underline{E}(A_i) \geq \overline{E}(A_r) \quad (21)$$

$$A_i \supseteq A_r \Leftrightarrow \underline{E}(A_i) \geq \underline{E}(A_r) \wedge \overline{E}(A_i) \geq \overline{E}(A_r) \quad (22)$$

\supseteq is called the strong dominance (or interval dominance) relation and \supseteq is called the interval bound dominance relation.⁵² Clearly the choice set defined from \supseteq will be a subset of that defined from \supseteq . My criticism of both these views runs along similar lines to my criticism of certain earlier views and reaches the same conclusion: the views do not cohere with FGI. Recall the version of the cup game in Table 3. My claim was that a decision method is not according with FGI unless at some point in the process of lowering x and k (but keeping them greater than 0) and raising y , the method deems bet 2 or 3 acceptable. But FGI actually demands more than this: furthermore, a decision method must at some point in this process (not necessarily the same point) stop deeming bet 1 acceptable. Otherwise, the method is not appropriately weighting your desires—your increasingly strong desire for the prize of bets 2 and 3 and increasingly weak desire for the prize of bet 1—by your degrees of belief—your increasingly strong degree of belief that bet 1 will not pay off and one of bets 2 or 3 will. But neither of the decision methods currently under consideration ever stops deeming bet 1 choiceworthy (not even, for example, when k is one in a million, x is one thousandth of one cent, and y is one billion dollars). The lower expectations of bets 2 and 3 are always 0 while the lower expectation of bet 1 is always positive ($k \cdot x$), so the lower expectation of bet 2 (likewise bet 3) can never be greater than or equal to the lower expectation of bet 1, and a fortiori can never be greater than or equal to the upper expectation of bet 1. Hence bet 2 (likewise bet 3) can never stand in either relation \supseteq or relation \supseteq to bet 1, and a fortiori bet 1 can never be dominated by bet 2 (likewise bet 3) relative to either of these relations. So bet 1 is always in the choice set, on both views currently under consideration. Thus, both views still allow you to pick bet 1, even when you are arbitrarily close to certain that bet 1 will not pay off—and even if it does you will win an arbitrarily small amount of money—and you are arbitrarily close to certain that one of bets 2 or 3 will pay off an arbitrarily large amount of money. No increase in your confidence, or decrease in payout x , or increase in payout y can possibly change this: and that contravenes FGI. (My view does stop deeming bet 1 acceptable at some point. You have to distribute the $1 - k$ belief mass applied to $\{C_2, C_3\}$ over $\{C_2\}$ and $\{C_3\}$. If you give more to the former, bet 2 will be preferable to bet 3—and vice versa; if you distribute the mass evenly, bets 2 and 3 will have the same expected value. But you cannot throw any of the mass away, and *however* you distribute it, at least one of bets 2 and 3 will be preferable to bet 1—provided x and k are low enough and y is high enough.⁵³ For higher values of k , or lower values of y that

⁵² It is important to distinguish the use of the term ‘dominance’ in the names of these relations from its use in the recipe described above for getting from some relation (regardless of whether ‘dominance’ appears in its name) to a choice set comprising the acts that are *nondominated* with respect to that relation.

⁵³ For a concrete example, suppose that $k = 0.01$, $x = 1$ and $y = 10$. Then bet 1’s expected payoff is one cent. Bet 2’s expected payoff could be anywhere from 0 to \$9.90, and similarly for bet 3—but they cannot vary independently: there is no way for *both* of them to be less than one cent.

bring y closer to x , bet 1 *will* be acceptable—or even mandated—on my view.⁵⁴ That is exactly the kind of sensitivity to belief and desire that FGI requires.)

The third view that fits the present pattern is the maximality view. Let the set of all probability functions compatible with your belief function be \mathfrak{P} and let the expected utility (in the classical sense) of act A_i relative to probability function P be $E_P(A_i)$. Now define the following relation between acts:

$$A_i \succeq A_r \Leftrightarrow \forall P \in \mathfrak{P} [E_P(A_i) \geq E_P(A_r)] \quad (23)$$

and then (following the pattern under consideration) define the choice set as the set of all nondominated acts with respect to this relation.⁵⁵ Again my criticism of this view is that it does not cohere with FGI, as can be seen by considering the cup game just discussed, where the view never stops deeming bet 1 acceptable. Bet 2 has a lower expected value than bet 1 relative to the probability function that assigns all of the $1 - k$ belief mass to cup 3, and bet 3 has a lower expected value than bet 1 relative to the probability function that assigns all of the $1 - k$ belief mass to cup 2—so bet 1 is nondominated and hence choiceworthy on this view.

The next view to be considered—the E -admissibility view—defines a choice set directly, rather than going via a relation between acts. Where \mathfrak{A} is the set of all available acts, define the choice set $\mathfrak{A}' \subseteq \mathfrak{A}$ as follows:

$$A_i \in \mathfrak{A}' \Leftrightarrow \exists P \in \mathfrak{P} \forall A_r \in \mathfrak{A} [E_P(A_i) \geq E_P(A_r)] \quad (24)$$

In other words, A_i is choiceworthy iff there exists a compatible probability function relative to which A_i maximises expected utility.⁵⁶ Note that an act will be in the choice set just defined if and only if it could be recommended by my view—by first picking a proportioning and then maximising expected value. So I have no objection to this view along the lines of my objections to the views just considered. But note that my view and the E -admissibility view are not the same. Consider for example bets 2 and 3 in the version of the cup game in Table 3 (where x and k are low enough and y high enough to rule out bet 1, on my view and the E -admissibility view). On the E -admissibility view they are both in the choice set and so the only rationally admissible attitude towards them is indifference: if the agent has to select a single bet she will just have to pick one of them. On my view however the agent is allowed to pick a proportioning relative to which bet 2 (say) has a higher expected value than bet 3—and if she does so then she will prefer bet 2 to bet 3. Thus on my view an agent is allowed to prefer bet 2 to bet 3, whereas on the E -admissibility view this is not allowed. This seems to me the correct position, given FGI. Your beliefs are *completely silent* on which of cups 2 or 3 specifically the ball will be under. This is quite different from a situation in which your degrees of belief are split evenly

⁵⁴ For a concrete example, suppose that $k = 0.6$, $x = 10$ and $y = 11$. Then bet 1's expected payoff is \$6. Bet 2's expected payoff could be anywhere from 0 to \$4.40, and similarly for bet 3—but however you divide up the 0.4 degree of belief applied to $\{C_2, C_3\}$, bet 1 will have a higher expected payoff than both bets 2 and 3.

⁵⁵ See Bewley (2002) and Walley (1991, Sect. 3.9).

⁵⁶ See Levi (1980, 96).

between cups 2 and 3. In the former situation, you are not failing to weight your desires by your beliefs if you break up your belief mass unevenly between cups 2 and 3. As I have argued, *any* way of breaking up your belief mass is acceptable from the point of view of FGI: you can only *pick* one. And once you pick one, it could be that one or other of bets 2 and 3 becomes preferable to you. A second advantage of my two-part view (first pick a proportioning, and then maximise expected value) over the one-part *E*-admissibility view (which simply defines a choice set) is that it naturally makes room for systematisation in local contexts in which consistency is a desideratum: as already discussed, the initial free picking step can (if desired, in a certain kind of setting) be locked into place with a mandated choice of a particular proportioning.

In Sect. 3, I presented a way of directly bringing together beliefs (modelled by belief functions) and desires (modelled in the standard way by utility functions) in order to (try to) implement FGI. I then criticised this view (arguing that it in fact fails to implement FGI) and in Sect. 4 introduced proportionings as a way of getting beliefs and desires to mesh, as a precursor to maximising expected value. The final view I want to consider involves a different way of bringing beliefs and desires together directly. Given a utility function U , each act A_i induces a real-valued function u_i which maps each S_j in the frame of discernment $\{S_1, \dots, S_n\}$ to $U(O_{i,j})$ (i.e. to the utility assigned to the outcome $O_{i,j}$ of act A_i in state S_j). Relative to an act A_i , we can order the states in the frame of discernment as S_1^i, \dots, S_n^i in such a way that $u_i(S_1^i) \leq \dots \leq u_i(S_n^i)$.⁵⁷ We also define $u_i(S_0^i)$ as 0 (so as to be able to write (25) below, rather than a lengthier expression). The proposal is that we assign act A_i an expected value $\text{exp}(A_i)$ equal to the Choquet integral of u_i with respect to the belief function Bel (and then pick an act with maximal expected value)⁵⁸:

$$\text{exp}(A_i) = \int_{Ch} u_i dBel = \sum_{j=1}^n (u_i(S_j^i) - u_i(S_{j-1}^i)) Bel(\{S_j^i, \dots, S_n^i\}) \quad (25)$$

We may think of what is going on here is as follows. We have an act A_i and we want to calculate an expected value for it. (Given A_i , we may for simplicity of presentation refer to the utility of the outcome of A_i relative to a certain state as the utility of that state.) We do this by summing a sequence of things—by, as it were, adding numbers into a pot. To get this sequence, we first order the states by their utilities, in a non-decreasing order. Now the first thing we throw into the pot is the utility of the first state—because we are bound to get at least that much utility (if we perform act A_i).⁵⁹ Then we look at the *increment* to the utility of the next state, S_2^i . We will gain that extra bit of utility (and maybe more, but that will be accounted for later in the process) if one of states S_2^i, \dots, S_n^i is actual, so we multiply the increment by our degree of belief in that set of states, and add the result to the pot. (In case the

⁵⁷ If there are multiple ways of doing this—because multiple states lead to outcomes with the same utility—which one we choose makes no difference to the expected value defined below.

⁵⁸ See Gilboa (1987) and Schmeidler (1989).

⁵⁹ Note that the first summand in (25) is simply $u_i(S_1^i)$ because we set $u_i(S_0^i) = 0$, and $\{S_1^i, \dots, S_n^i\}$ is the entire frame of discernment so $Bel(\{S_1^i, \dots, S_n^i\}) = 1$.

increment is in fact zero, we are adding nothing; cf. n.57.) Then we again look at the increment to the utility of the next state, S_3^i . We will gain *that* extra bit of utility (which is the first bit of the ‘maybe more’ mentioned above, for which we are now starting to account, as promised) if one of states S_3^i, \dots, S_n^i is actual, so we multiply the increment by our degree of belief in that set of states, and add the result to the pot—and so on until we have come to the last state. Now it turns out that $\exp(A_i) = \underline{E}(A_i)$ (Gilboa & Schmeidler, 1994, 51). Therefore, from my point of view, the present proposal suffers from the same flaws as maximin.

6 Precise probabilities redux

On one way of presenting the decision method proposed in Sect. 4, the way to make a decision when one’s degrees of belief are represented by a belief function is to pick a proportioning, which generates a probability function compatible with one’s belief function, and then maximise expected utility. The probability function does not represent one’s degrees of belief—the belief function still does that—but one acts as if one has probabilistic degrees of belief and is maximising expected utility. The objection might therefore arise: why the pointless detour through belief functions? If the end result is that one acts as if one has probabilistic degrees of belief and maximises expected utility, surely we should just adopt the traditional model in which your degrees of belief *are* probabilistic and you maximise expected utility.⁶⁰ If the probability is all that matters when it comes to action, shouldn’t we just drop the belief function from the picture?

We should not: the move via belief functions is not a pointless detour. First, on my view, the probability function is an inessential theoretical construct. The core point—FGI—is to decide in light of one’s beliefs and desires. The probability function is one way of linking evidence-based beliefs residing at the level of non-singleton sets of states, with utilities residing at the level of particular outcomes. There is an alternative way of forging this link—pooling utilities—that does not involve a probability function at all.

Second, there is the crucial conceptual point that we cannot say that the probability function represents your degrees of belief because degrees of belief need to respect the evidence and the probability function does not do so. The belief function that models your degrees of belief is non-Bayesian precisely because that is what the evidence warrants. The belief function cannot be dropped because without it, the picture would lack anything representing *degrees of belief* (which must respect the evidence).

Still, the objector may press, that’s all very well, but if, in the final analysis, you *behave as if* you have probabilistic degrees of belief and maximise expected utility, what content can we really give to the claim that your degrees of belief are not probabilistic? This brings us to my third point which is that the behaviour of a person *A* whose degrees of belief are modelled by a belief function (and who acts by picking a probability and then maximising expected utility) will *not* in general be the same as

⁶⁰ Elga (2010, 6) and White (2010, 180, n.19) argue in this sort of way against certain proposals about decision-making with imprecise probabilities.

the behaviour of a person B whose degrees of belief are probabilistic (and who acts by maximising expected utility). Consider the following case from White (2010, 175)⁶¹:

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 '.

Here's how the belief function model handles this case. The frame of discernment is $\{ph, pt, \bar{p}h, \bar{p}t\}$.⁶² 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. $\{pt, \bar{p}t\}$). Now 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 are now two left open ($ph, \bar{p}t$). So observing the coin land with ' p ' showing gives you conclusive evidence 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 gives a basic mass assignment m_3 that assigns 0.5 to $\{ph\}$ and 0.5 to $\{\bar{p}t\}$. m_3 —like the initial m_1 —generates a belief function that assigns 0.5 to h and 0.5 to t .⁶³

Now suppose we are asked to consider two bets: first a bet on p before the toss, and second a bet on h after the toss has been observed to come up ' p '. Prior to the toss, as just discussed, agent A has a belief function determined by a basic mass assignment of 0.5 to $\{ph, \bar{p}h\}$ (i.e. heads) and 0.5 to $\{pt, \bar{p}t\}$ (i.e. tails). From this, for purposes of deciding what to do in relation to the first bet, she generates a probability function by distributing the first 0.5 between $\{ph\}$ and $\{\bar{p}h\}$ and the second 0.5 between $\{pt\}$ and $\{\bar{p}t\}$. The probability assigned to p will then be the sum of whatever gets assigned to $\{ph\}$ and whatever gets assigned to $\{pt\}$. This could be anything between 0 and 1. Suppose for the sake of example (nothing depends on this) that A distributes the first 0.5 so that 0.15 goes to $\{ph\}$ and 0.35 goes to $\{\bar{p}h\}$, and the second 0.5 so that 0.15 goes to $\{pt\}$ and 0.35 goes to $\{\bar{p}t\}$. Her probability for p is then 0.3. Now suppose that B is an agent with probabilistic degrees of belief and that the probability function that represents his beliefs is exactly the one that A just generated. So B 's probability for p is 0.3. A and B act in the same way in relation to the first bet (they set the same betting ratio; or decline to pay 50¢ for a bet that pays out \$1 if p ; etc.). Now the coin is tossed and seen to come up showing ' p '. As just

⁶¹ The original lacks a negation symbol but that is clearly a typo so I have added one.

⁶² 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.

⁶³ Note that there is thus no loss of determinacy in your degrees of belief in heads and tails. This contrasts with the treatment of this case using imprecise probabilities which, notoriously, involves *dilation*. For further discussion see Smith (2022, §8).

discussed, A 's degree of belief in 'heads' remains 0.5—whereas B (after conditioning on $\{ph, \bar{p}t\}$) now assigns 'heads' probability 0.3. So A and B behave differently with regards to the second bet (they set different betting ratios; or A agrees to pay 40¢ for a bet that pays out \$1 if h but B declines; etc.).

What is going on here is that the key conceptual point made above does in fact manifest in behaviour. A carries around a belief function that represents her degrees of belief and *this* is what she updates when new evidence comes in. (She updates by combining her original basic mass assignment with the basic mass assignment warranted by the new evidence, using Dempster's rule.) She generates a probability from her belief function when she needs one for decision purposes but the belief function is still there representing her degrees of belief. The belief function is what gets updated in light of new evidence—and then a new probability is generated when she next needs to act. The old probability is not directly updated in light of evidence.⁶⁴ B , by contrast, has only the probability function to work with. It is being forced to do double duty of determining action (via the calculation of expected utility) and responding to evidence. So B updates the probability function when he sees the coin land with ' p ' showing—and this causes him to act differently from A with regards to the second bet.

The point can actually be pressed further: A is not only different from B but is in fact more reasonable. Given the setup of the case, the coin's landing ' p ' does not provide any evidence to shift one's view of h . If one was sure that p (or not p), then seeing ' p ' come up would tell one for sure that the coin landed heads (or tails). If one thought that p was quite/very un/likely, then seeing ' p ' come up would tell one that heads is quite/very un/likely (and so on). But the whole point is that one has no evidence at all about p : not that it is very or a little bit likely or unlikely or anything else. And in that situation, seeing the coin land ' p ' tells one nothing about whether it landed heads. This is exactly the result that the belief function story gives. Whereas B , who has only the probability function to work with, 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. And so B cannot avoid taking the observation of ' p ' as evidence for or against h . (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 .)⁶⁵

7 Criteria of mass assignments

In Sect. 2, I presented a picture in which a given body of evidence warrants a basic mass assignment which then determines (by Eq. (2)) a belief function which models degrees of belief. There is a debate in epistemology over whether (a) a given body of

⁶⁴ This might lead someone to object that my approach leaves the agent open to diachronic Dutch book: see Sect. 9.

⁶⁵ Furthermore, this is the root cause of the dilation problem for imprecise probabilities: it is because each member of the credal committee (that, according to imprecise probabilists, represents the agent's belief state) does what B does that the set of probabilities for h dilates from 0.5 to the entire unit interval. See n.63.

evidence warrants a particular doxastic attitude towards any proposition (the uniqueness view), or (b) different fully rational agents might adopt distinct doxastic attitudes towards a given proposition, on the basis of the same evidence (the permissive view).⁶⁶ For purposes of this paper I shall remain neutral on this question—but I shall address a prior issue: what does it *mean* for an agent to assign mass 0.3 (say), rather than some other mass such as 0.2, to a particular set of outcomes?⁶⁷ Shafer (1976, 20) writes:

I merely suppose that an individual can make a judgment. Having surveyed the sometimes vague and sometimes confused perception and understanding that constitutes a given body of evidence, he can announce a number that represents the degree to which he judges that evidence to support a given proposition and, hence, the degree of belief he wishes to accord the proposition

Shafer goes on explicitly to deny that the number is objectively determined by the evidence (in the terms above, he advocates permissiveness). He also distances himself from the views of those Bayesians who “have followed Frank Plumpton Ramsey and Bruno de Finetti in choosing to analyze degrees of belief as psychological facts, facts which can be discovered by observing an individual’s preferences among bets or risks but which may not bear any particular relation to any particular evidence” (p.21). This however means that Shafer leaves us in the dark about the significance, if any, of the agent’s choice of numbers: about what difference it would make if the agent announced 0.3 instead of 0.4—and hence about what these numbers really mean.⁶⁸

I face no such problem because when it comes to degrees of belief I do not take connections to *evidence* and connections to *preferences among bets* to be an either/or matter. As I argue elsewhere, degrees of belief have *both* kinds of connection. So we can indeed get a handle on what it means for a detective to assign (say) mass 0.3 rather than 0.4 to the set of male suspects (with the remainder being assigned to the frame of discernment) on the basis of a given body of evidence. It means (for example, among other things) that she may regard it as rationally permissible to pay 65¢ for a bet that pays \$1 if the perpetrator turns out not to be male.⁶⁹ In general, the connection between degrees of belief and decision-making described in Sect. 4 means that differences in mass assignments show up as differences in betting behaviour. Thus, on the belief function model advocated in this paper, while insisting that degrees of belief are constrained by evidence, we can also retain the traditional idea of understanding strength of belief by reference to betting behaviour. The precise *nature* of the connection is different—e.g. assigning mass 0.3 to the set of

⁶⁶ See e.g. Feldman (2007) and White (2005).

⁶⁷ An answer to this question could force a stance on the uniqueness issue—but my answer will leave the issue open.

⁶⁸ To be accurate, I should note that Shafer does address this issue in a later work where he notes that the lack of a “betting interpretation” of belief functions raises the question of what such degrees of belief mean (Shafer, 1981, 1–2) and goes on to offer an answer to the question.

⁶⁹ When generating a probability function for purposes of considering this bet, the maximum probability she can assign to ‘not male’ is 0.7 (allowing her to regard the bet favourably). It would be only 0.6 if she had assigned 0.4 to the set of males.

male suspects no longer equates to setting a betting ratio of precisely 0.3 on a bet that the perpetrator is male—but connections remain, and give us a concrete handle on the difference between distinct mass assignments.

8 Decisions under uncertainty vs risk

The view proposed in this paper yields a clear and unified picture of decision theory. On the traditional probabilistic picture, there is a distinction between decisions under risk—where there is a probability function that models the agent’s degrees of belief—and decisions under uncertainty—where there is no such probability function, because the agent has so little information that not only does he not know what will happen, he cannot even assign probabilities to the various possibilities. In the present picture, both kinds of situation are encompassed under the one umbrella and handled by the same decision theory. Traditional decisions under risk are cases where the belief function is Bayesian—i.e. a probability function. Traditional decisions under uncertainty are cases where the agent has so little information that the only focal set is the entire frame of discernment (which is assigned 1). In between there are intermediate cases where the focal sets are neither all singletons nor identical to the frame of discernment. All these cases are covered by the decision theory presented in Sect. 4. On the present picture, then, there are not two fundamentally different kinds of decision situation (risk and uncertainty). There is only one kind of case—but within it there are gradations, as the focal sets vary in size between singletons and the entire frame.

9 Sequential decisions

An agent whose evidence warrants degrees of belief represented by a non-Bayesian belief function and who follows the decision method proposed in Sect. 4 will not be subject to Dutch book—because in any given (synchronic) betting situation she will be acting as if she has probabilistic degrees of belief. For all that has been said, however, she could face a diachronic Dutch book. In this section, I consider diachronic Dutch books—and another kind of sequential decision problem, which has been argued by Elga (2010) to pose difficulties for imprecise probability views. While Elga did not consider belief functions, his case is relevant to my view too.

First let’s consider diachronic Dutch books. If an agent has degrees of belief represented by a non-Bayesian belief function Bel_1 at time t_1 and picks a compatible probability function P_1 for purposes of selecting bets, then learns E and updates (by Dempster’s rule) to a non-Bayesian belief function Bel_2 at time t_2 and then picks a compatible probability function P_2 for purposes of selecting bets, P_2 need not be the conditionalisation of P_1 on E . For example, suppose the sample space is $\Omega = \{S_1, S_2, S_3, S_4\}$. Let Bel_1 assign 0.5 each to $\{S_1, S_2\}$ and Ω . Let P_1 assign 0.3 each to $\{S_1\}$ and $\{S_4\}$ and 0.2 each to $\{S_2\}$ and $\{S_3\}$. Let E be—or warrant an assignment of 1 to— $\{S_2, S_4\}$. Then Bel_2 assigns 0.5 each to $\{S_2\}$ and $\{S_2, S_4\}$. Now let P_2 assign 0.5 each to $\{S_2\}$ and $\{S_4\}$. Then P_2 is not the same as the conditionalisation of P_1 on E , for the latter assigns 0.4 to $\{S_2\}$ and 0.6 to $\{S_4\}$. This

does not however mean that the agent is open to a diachronic Dutch book. Although it is often claimed in conversation that failure to update by conditionalisation leaves one open to Dutch book, in fact—as careful presentations in the literature make clear—susceptibility to diachronic Dutch book requires not merely that P_2 isn't the conditionalisation of P_1 on E . It requires furthermore that the bookie knows at t_1 what P_2 will be. Only then can the bookie know whether to buy or sell a certain bet at t_1 —in order to sell or buy it back (as the case may be) at t_2 —and thus inflict a sure loss.⁷⁰ So if you pick a probability on the fly when you need to act, you will not be susceptible to a diachronic Dutch book.

Earlier I said that it is allowable and perhaps desirable in certain contexts to bolt on to my view a fixed selection method—for example the TBM's Laplacean method—rather than picking a probability when one faces a decision. Now if one does do this—and makes it known that one has done so—then the point made in the previous paragraph would seem no longer to apply. Furthermore, the Laplacean method in particular can generate a later probability that is not the conditionalisation, on the intervening evidence, of the probability it generated earlier.⁷¹ None of this poses a problem for the views of this paper. The point simply becomes this: while it might be convenient in certain contexts to adopt the Laplacean method, contexts in which there are diachronic Dutch bookies lurking around are probably not among them (unless the benefit, in the context, of adopting the method outweighs the risk of losing money to a bookie).⁷²

Now let's consider Elga's scenario, which involves a proposition H —regarding which you lack the kind of evidence that would warrant *probabilistic* degrees of belief—and two bets: bet A sees you lose \$10 if H is true and win \$15 otherwise; bet B sees you win \$15 if H is true and lose \$10 otherwise. Your choices regarding these two bets can be depicted as in Fig. 3. Here's how to interpret this sort of diagram. After a node representing an opportunity which the agent may accept or reject (i.e. bet A, bet B), the up track represents acceptance and the down track represents rejection. After a node containing a proposition (i.e. H), the up track represents the proposition's being true and the down track represents its being false. Nodes containing numbers represent outcomes (dollar amounts). So looking at the top part of Fig. 3, the idea is that if you accept bet A, then if H is true you lose \$10 and if H is false you gain \$15; while if you reject bet A, then you gain (and lose) nothing (the

⁷⁰ The classic presentation of the diachronic Dutch book argument for conditionalisation is Teller (1973)/ (1976). Note in particular: “if at time 0 *the agent knows...* what his new belief function... will be... then a bookie *who knows no more nor less* than the agent can induce the agent to buy and sell bets on which he will have a net loss whatever happens. ...No *explicitly formulated plan* for changing beliefs in the face of new evidence is reasonable unless... the plan calls for changing beliefs by conditionalization” (p. 223 1973/p. 210 1976, my emphases) and “In this case, the bookie cannot inflict sure loss because he cannot tell for sure whether he should buy or sell the initial bets” (p. 225/p.212).

⁷¹ See Smets (1993, Sect. 2.2) for an example in which this occurs. Note that Smets argues that the diachronic Dutch book argument in fact *fails* against the TBM.

⁷² Compare: if you have a choice of ways of walking home from the station, you can pick one each time, or you can instead lay out a timetable: route A on Monday, B on Tuesday, etc. Either approach is fine in general—but if you live in a high crime area then it might be better to randomise your movements than to follow a fixed plan.

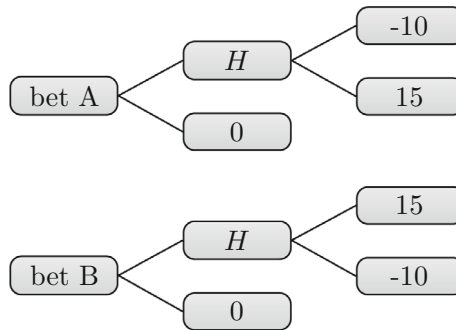


Fig. 3 Bets A and B: version I

status quo is maintained). Similarly for the bottom part of the diagram and for further such diagrams below.

Elga is going to offer you bet A and then—so soon afterwards that you will not have lost, gained or reconsidered any evidence—offer you B. You know in advance that this will happen. Elga claims:

Any perfectly rational agent who is sequentially offered bets A and B in the above circumstances (full disclosure in advance about the whole setup, no change of belief in H during the whole process, utilities linear in dollars) will accept at least one of the bets. (Elga, 2010, 4)

On the view proposed in this paper, however, you could end up rejecting both bets. Suppose that your degrees of belief about H are represented by a non-Bayesian belief function generated by a mass assignment of x to $\{H\}$, y to $\{\neg H\}$, and $1 - x - y$ to $\{H, \neg H\}$. To decide whether or not to take bet A, you divide the $1 - x - y$ assigned to $\{H, \neg H\}$ between $\{H\}$ and $\{\neg H\}$ —generating a probability function—and then maximise expected utility. If $\{H\}$ gets a probability more than 0.6, then the expected utility of accepting bet A will be less than zero—which is the expected utility of rejecting the bet—and so you will reject it. To decide whether or not to take bet B, you proceed in the same kind of way. If $\{H\}$ gets a probability less than 0.4, then the expected utility of accepting bet B will be less than zero—which is the expected utility of rejecting the bet—and so you will reject it. So on my view, you are rationally permitted to reject bet A and then reject bet B. Note that this will not happen if you use the same proportioning of $\{H, \neg H\}$ for both decisions.⁷³ However I have already argued that while it might make sense to replace the picking part of my decision process with a fixed selection mechanism in certain contexts for pragmatic reasons, this is *not* a requirement of rationality.

My basic position on Elga's scenario is that (for all the reasons given above in support of my proposal) my view gives the correct result—and that Elga's claim that

⁷³ It is part of the setup that your belief in H stays the same—i.e. in my terms, the belief function representing your degrees of belief does not change. If you divide up your degree of belief in $\{H, \neg H\}$ in the same way both times, then $\{H\}$ gets the same probability both times. Hence, it does not get probability more than 0.6 one time and less than 0.4 the other.

a perfectly rational agent may not end up having rejected both bets is incorrect. Other authors agree that Elga is incorrect.⁷⁴ So we could leave the matter here. However taking the discussion a couple of steps further will shed light on how an agent would employ the decision theory presented in this paper in scenarios in which future decisions need to be factored into present ones.

The first point to discuss is that the above account of how one would decide in Elga's case would be exactly the same if one were offered bet A with no knowledge that bet B was coming—and then later happened to be offered bet B. Yet Elga stresses that full disclosure in advance is an important part of the setup of his case. He does not think an agent is irrational if she does not know bet B is coming when she rejects bet A—and then, when bet B is presented, goes on to reject it. Rather, he claims that a perfectly rational agent who knows both bets are coming—in quick enough succession that she won't get any new evidence relevant to H after bet A and before bet B—will accept at least one of the bets. Elga claims:

The bets are great because if you accept both of them, you'll be sure to win \$15 on one, lose \$10 on the other, and so gain \$5 overall. In other words, bets A and B together guarantee a sure gain, just as a Dutch Book guarantees a sure loss. (Elga, 2010, 4)

To build the idea that when you consider bet A, you also know bet B is coming, we can depict the situation as in Fig. 4. Note that the lower version of bet B is the same as in Fig. 3—but the upper version is different. The point here is that if you have accepted bet A, then the overall outcome of accepting, or rejecting, bet B will be different from what it would have been had you rejected bet A: now your final position, when the truth or falsity of H is revealed and the bets are settled, will be the sum of your payouts on bets A and B. As you have full disclosure in advance, you can already envisage all this at the point of considering bet A.

Even with this more complex setup, the view proposed in this paper allows a rational agent to reject A and then go on to reject B (and this is the right result). Suppose you are standing at the bet A node, deciding whether to go up or down. In the distance, you see the possible outcomes. In between, you see your future decision on bet B, and the question of H . Your degrees of belief concerning how likely you are to get the various payoffs therefore depend on your degrees of belief concerning H , and concerning what you will do when faced with bet B. So let's consider the latter. (The idea is that this sort of consideration is precisely what you yourself engage in, while deliberating over bet A.) Suppose you are weighting up bet B. At that point, the only thing standing between your decision and the outcomes is the question of H . So your decision on B will depend on your degrees of belief (at that point) concerning H . More specifically, you will pick a proportioning of $\{H, \neg H\}$. This will generate a probability p for H . You then calculate the expected utility of accepting and rejecting. At the top version of bet B, these expected utilities are (respectively) 5 and $p \cdot -10 + (1 - p)15 = 15 - 25p$. You will accept if the former exceeds the latter, i.e. if $p > 0.4$. At the bottom version of bet B, the expected utilities are $p \cdot 15 + (1 - p) \cdot -10 =$

⁷⁴ For a compelling rebuttal of Elga, see Bradley and Steele (2014a); also e.g. see Hedden (2015) and cf. Chandler (2014, 1281–2). Earlier, Levi (1985, §3) had (as Elga notes) anticipated this sort of case and argued against the kind of position that Elga adopts.

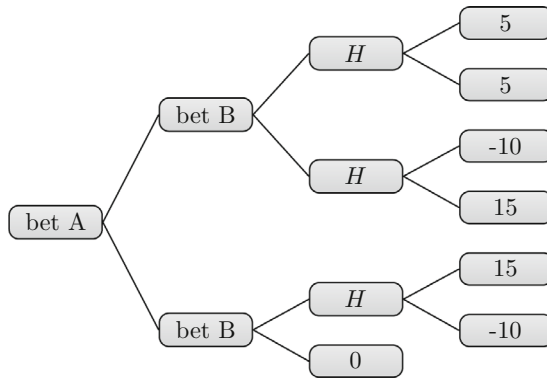


Fig. 4 Bets A and B: version II

$25p - 10$ and 0 . You will accept if the former exceeds the latter, i.e. if $p > 0.4$. Now return to the bet A node. As we saw, your degrees of belief concerning whether you will end up with each of the possible payoffs depend on your degrees of belief concerning H , and concerning what you will do when faced with bet B. Regarding the latter, you’ve just worked out that (whether you accept or reject bet A) you’ll accept bet B if, at the point of considering bet B, you assign H a probability greater than 0.4 . The view proposed in this paper says that at bet B, you assign H a probability by picking a proportioning of $\{H, \neg H\}$. This (in itself) determines *absolutely nothing* about what beliefs you should have in advance about whether you will pick this or that proportioning—and hence assign H this or that probability. Your beliefs about this will depend (as always) on your evidence. Maybe you have noticed that in the past you tend to make certain kinds of pickings—but maybe you have discerned no pattern. Or maybe you have adopted the Laplacean method of picking a proportioning and feel certain that you will continue to use it in the future. All of this is up for grabs. Now however you form your degrees of belief at the time of considering bet A, and however you proportion focal sets to arrive at probabilistic degrees of belief, you will assign probabilities to each of the four cells in the partition generated by the two propositions H and ‘when considering bet B, you will assign H a probability greater than 0.4 ’ (BPH for short)⁷⁵:

BPH	H	Probability
1	1	a
1	0	b
0	1	c
0	0	$1 - a - b - c$

⁷⁵ In the table, 1’s and 0’s are truth values. So the first cell in the partition is the one where BPH and H are both true—and so on.

So now you can assign the following expected utility to the act of accepting bet A⁷⁶:

$$a.5 + b.5 + c.-10 + (1 - a - b - c)15 = 15 - 10a - 10b - 25c \quad (26)$$

and the following expected utility to the act of rejecting bet A:

$$a.15 + b.-10 = 15a - 10b \quad (27)$$

You will accept A iff (26) exceeds (27). A bit of calculation shows that this occurs just in case $a + c < 0.6$. This *could* fail to occur—in which case you will reject A. So even in this more complex setup, it is possible for your degrees of belief at the point of considering bet A—your degrees of belief about H , and about the probability you will assign H when it comes to considering bet B—to lead you to reject bet A. And then of course it is possible that when you come to consider bet B, you assign a probability to H that leads you to reject bet B. So you could end up having rejected bets A and B (and there is nothing irrational about this).

But what of Elga's claim that bets A and B together guarantee a sure gain? There is no point in Fig. 4 at which you can choose bets A and B together. You only get to choose bet A (or not)—and then bet B (or not). There is no choice point at which you turn down a sure gain in favour of a sure zero dollars—and no point at which you choose irrationally (even if you end up with zero dollars). However—and this is the second point to discuss—there is in fact a third, even more complex way of looking at the situation, in which you *can* at the outset choose a sure \$5. For you can see in advance that if you accept bet A and accept bet B, then you will get \$5. So—if you think of this as a real-world scenario—you do in fact have the option of bypassing the decision-making process altogether. You have the option of deciding at the outset that you will not even weight up the options when offered bet A, or when offered bet B—you will simply say 'I accept' on both occasions. We can depict the situation as in Fig. 5. At the outset you now have a new, initial choice: take the decision theory route (the down track, which from that point on is just the same as in Fig. 4); *or* bypass it (the up track). The idea of the up track is *not* that you somehow try to bind your decisions in advance: it is that you do not make any (further) decisions at all. Decision theory tells you how to make decisions. It does not tell you which real-world scenarios *are* decisions. You give decision theory a decision as input and it gives you as output a choice set. You do not give it as input a real-world three- or four-dimensional scene or scenario involving people with cash and so on. You can

⁷⁶ I am here making a harmless assumption, for the sake of simplicity of presentation. You know that you'll accept bet B if you assign H a probability greater than 0.4. Note 'if' here—not 'if and only if'. You can see that if you assign H probability 0.4, the expected utilities of accepting and rejecting bet B will be the same, and you will *pick* a course of action (this is what happens when there are multiple options that all maximise expected utility). I'll now assume that (for whatever reason) you feel certain that if you are made an offer, and accepting and declining both maximise expected utility, then you will decline. This assumption turns the 'if' into 'if and only if'. The assumption need not be true in general, of all agents—but making it here simplifies the presentation, and has no impact on my argument (below) that *there exist* circumstances under which a rational agent may reject A and then reject B. It has no impact because there are certainly circumstances in which it is rational for a particular agent to be certain in this sort of way about what he will do if he has to pick one out of multiple options that all maximise expected utility.

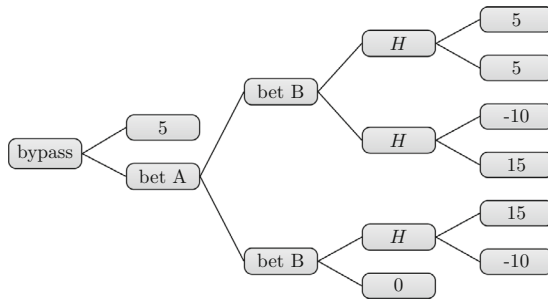


Fig. 5 Bets A and B: version III

treat such a scene as a decision—abstract away the choices you have, the outcomes you will get, and so on—and then feed these into your decision theory. My point now is that faced with Elga’s real-world scenario, you actually have a *prior* option which is not to treat it as a decision situation at all. (Compare: your theory of etiquette directs you not to put your feet on the table. But it does not tell you which real-world objects *are* tables. If you set up camp in the woods, then once you deem a certain flat rock to be the camp table and start using it as such, your theory of etiquette kicks in and directs everyone not to put their feet on it. But the theory of etiquette does not itself tell you that that rock must be treated as a table in the first place.) You could just view Elga’s scenario as a way of getting \$5. Saying ‘I accept’ to bet A and then saying ‘I accept’ to bet B would not then be *deciding* to accept bet A and then *deciding* to accept B: it would just be a procedure that leads to \$5. It would be analogous to the procedure of putting a bank card into the ATM, entering the PIN, and withdrawing \$5. On *this* way of approaching Elga’s real-world scenario, ‘accepting bet A’ would not be a decision but a step in a sequence of moves leading to \$5: analogous to pushing a certain button on the ATM—one of a sequence of buttons you need to push to get your \$5. And note that while you may decide to go to the ATM and withdraw \$5, typically you do not then decide whether or not to press each button—weighting up the pros and cons of pressing and not pressing. You could in certain (atypical) circumstances do this—but usually you do not. My point is that you *could* take a similar approach to bets A and B. (Of course you could decide on the up track and then find yourself getting sucked into the decision-making process later on—but this *need not* happen. You *could* successfully take the up track and make no further decisions.)

On this picture, the initial ‘bypass’ node *is* a decision. So the question arises, should you take the up track—bypassing any *further* use of decision theory and treating the bets as you would an ATM? Or should you take the down track—in which case you will then have two further decisions to make later? The decision theory proposed in this paper *allows* you (in certain possible circumstances) to take the down track. Even though the up track definitely leads to \$5, you might (depending on your beliefs) value the down track *higher* than this and hence choose to take it. To value the down track, you consider the possible outcomes at the end of it. In between you and these outcomes lie *three* things: the decision you will make on

A; the decision you will make on B; and the question of H . Whatever your beliefs are about these things (cf. the discussion above)—and however you proportion focal sets to arrive at probabilistic degrees of belief (for purposes of deciding whether to bypass)—you will assign probabilities to each of the eight cells in the partition generated by the three propositions H , BPH and ‘when considering bet A, you will assign probabilities in such a way that $a + c < 0.6$ ’ ($APBH$ for short):

$APBH$	BPH	H	Probability
1	1	1	d
1	1	0	e
1	0	1	f
1	0	0	g
0	1	1	h
0	1	0	i
0	0	1	j
0	0	0	$1 - d - e - f - g - h - i - j$

So now you can assign the following expected utility to the act of declining the bypass:

$$\begin{aligned} d.5 + e.5 + f.-10 + g.15 + h.15 + i.-10 \\ = 5d + 5e - 10f + 15g + 15h - 10i \end{aligned} \quad (28)$$

You will decline the bypass if (28) exceeds 5. Clearly that *could* occur (e.g. if $d = e = j = 0.125, f = i = 0.07$ and $g = h = 0.18$, then (28) equals 5.25). And then of course you could go on to reject bet A, and then to reject B. And this is the right result: given certain beliefs on your part about H and about the probabilities you will assign in the future (when you need to make decisions on A, and then on B), such behaviour is entirely rationally acceptable. In rejecting the bypass option, you are turning down \$5. But you are doing so in the expectation (relative, as always, to your own degrees of belief) of even greater gain. At every choice point, you take the option that looks better to you. Things may work out badly for you, but that does not mean you acted irrationally. (They might also work out well, and you end up with \$15.) The key point is that although you may turn down \$5, and later end up with nothing, you never turn down \$5 in favour of a sure \$0. We could add a *third* track to the initial node in Fig. 5, leading directly to \$0. (For future reference, call this the ‘middle track’—with the up and down tracks being as shown in Fig. 5.) For clearly, in just the way that saying ‘I accept’ to bet A and then saying ‘I accept’ to bet B is a procedure that leads to \$5, saying ‘no thank you’ to bet A and then saying ‘no thank you’ to bet B is a procedure that leads to \$0. However, the decision theory proposed in this paper would not allow you to take this middle track. Whatever your beliefs about what might happen on the down track, you value the middle track strictly lower than the up track. Thus you will always choose the up track or the down track—never the middle track. The decision theory proposed in this paper potentially (depending

on your beliefs) allows you to make choices that *could* result in \$0 in the end—even when you could at the outset make a choice that guarantees you \$5. It does not allow you to make a choice that *guarantees* you \$0 when you could make a choice that guarantees you \$5. All of this is just as it should be.⁷⁷

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflict of interest The author declares that there are no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Antoine, A., & Nicole, P. (1996). *Logic or the art of thinking: Containing, besides common rules, several new observations appropriate for forming judgment*. Cambridge University Press, Cambridge. Translated and edited by Jill Vance Buroker.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.
- Bewley, T. F. (2002). Knightian decision theory. Part I. *Decisions in Economics and Finance*, 25:79–110. Originally circulated as Discussion Paper no. 807 of the Cowles Foundation at Yale University, November 1986.
- Bradley, S. (2019). A counterexample to three imprecise decision theories. *Theoria*, 85, 18–30.
- Bradley, S., & Steele, K. (2014). Should subjective probabilities be sharp? *Episteme*, 11(3), 277–89.
- Bradley, S., & Steele, K. (2014). Uncertainty, learning, and the “problem” of dilation. *Erkenntnis*, 79(6), 1287–1303.
- Chandler, J. (2014). Subjective probabilities need not be sharp. *Erkenntnis*, 79(6), 1273–86.
- Cobb, B. R., & Shenoy, P. P. (2003). A comparison of methods for transforming belief function models to probability models. In Thomas Dyhre Nielsen and Nevin Lianwen Zhang, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 7th European Conference, ECSQARU 2003*, pp. 255–66. Berlin: Springer.
- Cobb, B. R., & Shenoy, P. P. (2006). On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41, 314–30.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2), 325–39.
- Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B (Methodological)*, 30, 205–47.
- Deneuëx, T. (2019). Decision-making with belief functions: A review. *International Journal of Approximate Reasoning*, 109, 87–110.
- Destercke, S. (2010). A decision rule for imprecise probabilities based on pair-wise comparison of expectation bounds. In Christian Borgelt, Gil González-Rodríguez, Wolfgang Trutschnig, María Asunción Lubiano, María Ángeles Gil, Przemysław Grzegorzewski, and Olgierd Hryniewicz, editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77 of *Advances in Intelligent and Soft Computing*, pp. 189–97. Springer, Berlin.
- Dubois, D., & Prade, H. (1986). A set theoretic view of belief functions: Logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12, 193–226.


⁷⁷ Thanks to the editor, the anonymous referees and Michael Nielsen for very helpful comments.

- Elga, A. (2010). Subjective probabilities should be sharp. *Philosophers' Imprint*, 10(5), 1–11.
- Etner, J., Jeleva, M., & Tallon, J.-M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2), 234–70.
- Feldman, R. (2007). Reasonable religious disagreements. In Louise M. Antony, editor, *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, pp. 194–214. Oxford University Press.
- Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese*, 53, 361–86.
- Gilboa, I. (1987). Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics*, 16, 65–88.
- Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18, 141–53.
- Gilboa, I., & Schmeidler, D. (1994). Additive representations of non-additive measures and the Choquet integral. *Annals of Operations Research*, 52, 43–65.
- Halpern, J. Y. (2017). *Reasoning about uncertainty* (2nd ed.). Cambridge: MIT Press.
- Hedden, B. (2015). Options and diachronic tragedy. *Philosophy and Phenomenological Research*, 90(2), 423–51.
- Huntley, N., Hable, R., & Troffaes, M. C. M. (2014). Decision making. In Thomas Augustin, Frank P.A. Coolen, Gert de Cooman, and Matthias C.M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pp. 190–206. Wiley, Chichester.
- Hurwicz, L. (1951). The generalised Bayes-minimax principle: A criterion for decision-making under uncertainty. Cowles Commission Discussion Paper: Statistics No. 355.
- Jaffray, J.-Y. (1988). Application of linear utility theory to belief functions. In B. Bouchon, L. Saitta, & R. Yager (Eds.), *Uncertainty and intelligent systems* (pp. 1–8). Berlin: Springer.
- Jeffrey, R. (1983). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press.
- 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, 549–77.
- Kaplan, M. (1996). *Decision theory as philosophy*. Cambridge: Cambridge University Press.
- Kyburg, H. E., Jr. (1983). Rational belief. *The Behavioral and Brain Sciences*, 6(231–45).
- Kyburg, H. E., Jr., & Teng, C. M. (2001). *Uncertain inference*. Cambridge: Cambridge University Press.
- Laplace, P. S., Marquis de (1951). *A philosophical essay on probabilities*. Dover Publications, New York. Translated from the Sixth French Edition by Frederick Wilson Truscott and Frederick Lincoln Emory.
- 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*. Cambridge: MIT Press.
- Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Philosophy of Science*, 52(3), 390–409.
- Moss, S. (2015). Credal dilemmas. *Noûs*, 49(4), 665–83.
- Ramoni, M., & Sebastiani, P. (2001). Robust Bayes classifiers. *Artificial Intelligence*, 125, 209–26.
- Ramsey, F.P. (1931). Truth and probability (1926). In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays* (pp. 156–98). London: Routledge and Kegan Paul.
- Resnik, M. D. (1987). *Choices: An introduction to decision theory*. Minneapolis: University of Minnesota Press.
- Satia, J. K., & Lave, R. E., Jr. (1973). Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21, 728–40.
- Savage, L. J. (1972). *The foundations of statistics* (2nd ed.). New York: Dover Publications.
- Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57(3), 571–87.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.
- Shafer, G. (1979). Allocations of probability. *The Annals of Probability*, 7(5), 827–39.
- Shafer, G. (1981). Constructive probability. *Synthese*, 48, 1–60.
- Shapley, L. S. (1953). A value for n-person games. In Harold William Kuhn and Albert William Tucker, editors, *Contributions to the Theory of Games*, volume II, pages 307–17. Princeton University Press.
- Smets, P. (1981). The degree of belief in a fuzzy event. *Information Sciences*, 25, 1–19.
- Smets, P. (1988). Belief functions. In Philippe Smets, E.H. Mamdani, Didier Dubois, and Henri Prade, editors, *Non-Standard Logics for Automated Reasoning*, pages 253–77. London: Academic Press.

- Smets, P. (1993). No Dutch book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. In R. Scozzafava (Ed.), *SIS Workshop on Probabilistic Expert Systems* (pp. 181–204). Rome: Societa Italiana di Statistica.
- Smets, P. (2005). Decision making in the TBM: The necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38, 133–47.
- 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. (2022). Respecting evidence: Belief functions not imprecise probabilities. *Synthese*, 200(475), 1–30.
- Strat, T. M. (1990). Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4, 391–417.
- Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, 42(1), 139–165.
- Teller, P. (1973). Conditionalization and observation. *Synthese*, 26(2), 218–258.
- Teller, P. (1976). Conditionalization, observation, and change of preference. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 1, pages 205–259 Dordrecht: Reidel.
- Troffaes, M. C. M. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1), 17–29.
- Troffaes, M. C. M., & de Cooman, G. (2014). *Lower Previsions*. Chichester: Wiley.
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42), 230–265.
- Ullmann-Margalit, E., & Morgenbesser, S. (1977). Picking and choosing. *Social Research*, 44(4), 757–85.
- van Fraassen, B.C. (1989). *Laws and symmetry*. Oxford: Clarendon Press.
- van Fraassen, B.C. (1990). Figures in a probability landscape. In J. Michael Dunn and Anil Gupta, editors, *Truth or Consequences: Essays in Honor of Nuel Belnap*, pages 345–56. Kluwer Academic Publishers, Dordrecht.
- von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior* (3rd ed.). New York: John Wiley & Sons.
- Wald, A. (1945). Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, 46(2), 265–280.
- Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.
- Walley, P. (1996). Measures of uncertainty in expert systems. *Artificial Intelligence*, 83, 1–58.
- Walley, P. (2006). Imprecise probabilities. In Samuel Kotz, N. Balakrishnan, Campbell Read, and Brani Vidakovic, editors, *Encyclopedia of Statistical Sciences* (2nd ed.), volume 5, pp.3353–6. John Wiley & Sons, Hoboken NJ.
- Wasserman, L. A. (1990). Prior envelopes based on belief functions. *The Annals of Statistics*, 18(1), 454–64.
- White, R. (2005). Epistemic permissiveness. *Philosophical Perspectives*, 19, 445–59.
- White, R. (2010). Evidential symmetry and mushy credence. In Tamar Szabó Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology*, volume 3, pages 161–186. New York: Oxford University Press.
- Williams, P. M. (1982). Discussion of Professor Shafer's paper. *Journal of the Royal Statistical Society, Series B (Methodological)*, 44(3), 341–3.
- Yager, R. R. (1992). Decision making under Dempster–Shafer uncertainties. *International Journal of General Systems*, 20, 233–45.
- Yager, R. R. (2004). Decision making using minimization of regret. *International Journal of Approximate Reasoning*, 36, 109–28.
- Zaffalon, M., Wesnes, K., & Petrini, O. (2003). Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artificial Intelligence in Medicine*, 29, 61–79.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Nicholas J. J. Smith¹ 

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

¹ Philosophy, The University of Sydney, Sydney, NSW 2006, Australia