



Probability functions, belief functions and infinite regresses

David Atkinson¹ · Jeanne Peijnenburg¹ 

Received: 13 November 2019 / Accepted: 17 October 2020 / Published online: 7 November 2020
© The Author(s) 2020

Abstract

In a recent paper Ronald Meester and Timber Kerkvliet argue by example that infinite epistemic regresses have different solutions depending on whether they are analyzed with probability functions or with belief functions. Meester and Kerkvliet give two examples, each of which aims to show that an analysis based on belief functions yields a different numerical outcome for the agent's degree of rational belief than one based on probability functions. In the present paper we however show that the outcomes are the same. The only way in which probability functions and belief functions can yield different solutions for the agent's degree of belief is if they are applied to different examples, i.e. to different situations in which the agent finds himself.

Keywords Infinite epistemic regress · Probability function · Belief function · Uniqueness

1 Introduction

It is a truth widely acknowledged that a belief which is justified must be based on a reason. This reason is either justified or unjustified. If it is unjustified, it cannot really provide justification, but if it is justified, then it must be based on a second reason, which in turn is either justified or unjustified, and so on. Enter the notorious regress problem: (a belief in) a proposition E_0 is justified by (a belief in) E_1 , which is justified by (a belief in) E_2 , et cetera, *ad infinitum*. As a justificatory chain, such a regress not only seems at odds with our finite natures, it also appears to be inconsistent, for it blocks rather than brings justification. It is as if we are given a cheque with which we go to a bank teller, who gives us a new cheque with which we go to another bank teller,

✉ Jeanne Peijnenburg
jeanne.peijnenburg@rug.nl
David Atkinson
d.atkinson@rug.nl

¹ Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

and so on. Never do we encounter a teller who actually converts our cheque into bars of gold. Similarly, since the justification in an infinite epistemic regress is for ever postponed and never cashed out, a proposition which receives its justification from an infinite regress actually is *not* justified, for the justification never materializes.

The regress problem has plagued epistemology for at least two millennia. Several solutions have been proposed, but each has its own drawbacks.¹ The problem is no small beer, for the very existence of justified beliefs and thus of knowledge depends on it. Bonjour called considerations pertaining to the problem “perhaps the most crucial in the entire theory of knowledge” and Huemer speaks of “the most fundamental and important issues in all of human inquiry”.²

As it turns out, a great deal of the problem springs from the fact that the justification relation has been seen, sometimes explicitly but more often tacitly, as some sort of entailment, where E_k is implied by E_{k+1} for all k . In that case it is of course impossible to determine the truth value of E_0 if the chain is infinitely long. However, once the relation is given a probabilistic interpretation, in which E_k is merely made more probable by E_{k+1} , then the probability value of the target proposition, $P(E_0)$, can be determined even if the chain goes on forever.³ If E_0 is probabilistically justified by E_1 , which in turn is probabilistically justified by E_2 , and so on, then it can be shown that $P(E_0)$ is well-defined. It does not go to zero, as has been maintained by C.I. Lewis, nor does it remain forever unknown, as for example Nicholas Rescher thought.⁴ Rather it tends to a unique, positive value between zero and one. In this sense a probabilistic regress, as Frederik Herzberg has called it, is not inconsistent in the way that a non-probabilistic one is. Unlike the truth value of E_0 in a non-probabilistic regress, the probability value of E_0 in a probabilistic regress is well-defined, so justification can materialize in the latter but not in the former.⁵

The consistency of probabilistic regresses has been demonstrated on the basis of Kolmogorov probability functions.⁶ In a recent paper, Ronald Meester and Timber Kerkvliet have made a welcome and original attempt to analyze such regresses in terms of Shafer belief functions.⁷ Their findings reinforce the view that epistemic regresses are consistent once they have been given a probabilistic interpretation rather than an interpretation in terms of entailment. This is a notable result, for belief functions and probability functions are rather different. First, belief functions are weaker than probability functions, containing the latter as special cases. Second, belief functions have been hailed as being particularly useful for modelling lack of information,

¹ For recent overviews, see for example (Aikin 2011; Aikin and Peijnenburg 2014; Turri and Klein 2014).

² Bonjour (1985), p. 18; Huemer (2016), p. 16.

³ E_k is made more probable by E_{k+1} iff $P(E_k|E_{k+1}) > P(E_k|E_{k+1}^c)$, where E^c is the complement of proposition E , taken as a set of possible worlds. For the most part we will in this paper use set theory notation (rather than the notation for logical languages); for first, it is the notation used by Meester and Kerkvliet, and second, both Kolmogorov and Shafer defined their functions over algebras and not over logical languages. The inequality $P(E_k|E_{k+1}) > P(E_k|E_{k+1}^c)$ is a necessary but not sufficient condition for probabilistic justification. See Peijnenburg (2007) and Atkinson and Peijnenburg (2017).

⁴ Lewis (1929), pp. 327–328; Rescher (2010), pp. 36–37.

⁵ Herzberg (2010).

⁶ Peijnenburg (2007), Atkinson and Peijnenburg (2017), Kolmogorov (1933).

⁷ Meester and Kerkvliet (2019), Shafer (1976), Shafer (1981).

which probability functions can only handle by means of probability intervals.⁸ Third, whereas the value of the unconditional probability $P(E_0)$ is determined solely by the conditional probabilities, $P(E_k|E_{k+1})$ and $P(E_k|E_{k+1}^c)$, the value of the unconditional belief $Bel(E_0)$ cannot be expressed solely by the conditional beliefs $Bel(E_k|E_{k+1})$ and $Bel(E_k|E_{k+1}^c)$.⁹

Although probabilistic regresses turn out to be consistent in both analyses, Meester and Kerkvliet argue that an analysis with belief functions may yield *a different numerical outcome* than one using probability functions. In other words, given a particular infinite epistemic regress, the value of $Bel(E_0)$ may differ from that of $P(E_0)$. This would mean that our degree of rational belief in E_0 could vary, dependent on whether we analyze the regress with probability functions or with belief functions.

Meester and Kerkvliet argue for this claim by means of examples. In their Sect. 3, they construct two examples aimed at showing that Kolmogorov probability functions and Shafer functions give different outcomes for one and the same probabilistic regress. In Example 3.1 they find that the agent's degree of belief in E_0 is one half if one uses belief functions, $Bel(E_0) = \frac{1}{2}$, whereas with probability functions the degree of belief is a bit less: $P(E_0) = \frac{3}{7}$. In Example 3.2, Shafer belief functions again give $Bel(E_0) = \frac{1}{2}$, but a probability distribution yields $P(E_0) = \frac{1}{3}$.

Contrary to what the paper of Meester and Kerkvliet may suggest, it is however not so that $P(E_0) = \frac{3}{7}$ in Example 3.1, nor is it so that $P(E_0) = \frac{1}{3}$ in Example 3.2. A detailed analysis of Examples 3.1 and 3.2 shows that in both examples $P(E_0) = \frac{1}{2}$. So these examples do not show that a particular regress may have different solutions when analyzed with belief functions or with probability functions: in both examples, $Bel(E_0) = P(E_0) = \frac{1}{2}$. They only show that a regress may have different solutions *in different situations*; applied to the same situation or example, $Bel(E_0)$ and $P(E_0)$ are numerically identical.

Our paper is set up as follows. In Sect. 2 we call to mind the systems of Kolmogorov and Shafer, and the difference between probability functions and belief functions. In Sect. 3 we summarize Meester and Kerkvliet's reasoning about their Example 3.1, and we explain that a Kolmogorov analysis of this example gives $P(E_0) = \frac{1}{2}$. In Sect. 4 we do the same for their Example 3.2, showing that here, too, $P(E_0) = \frac{1}{2}$. Thus neither Example 3.1 nor Example 3.2 shows that a particular regress can have different solutions when applied to the same situation. Can there exist other examples which do the job? In Sect. 5 we explain that this is impossible: probability functions and belief functions always yield the same numerical outcome when applied to a particular probabilistic regress in a particular situation.

⁸ For a more sceptical view about the Shafer approach, see for instance (Pearl 1988).

⁹ For the former, see Peijnenburg (2007) and Atkinson and Peijnenburg (2017); for the latter, see Meester and Kerkvliet (2019), Sect. 4.

2 Probability functions and belief functions

In this section we recall the axioms of Kolmogorov and of Shafer, and we shall explain the basic similarities and differences.¹⁰

The probability functions are defined on a σ -algebra over a set Ω , that is a collection of Ω 's subsets which includes Ω , and is closed under complement and under countable unions. Originally these subsets are related to events, and we follow Kolmogorov in stating the axioms in this language. However, they can also refer to (beliefs in) propositions, and this is of course more natural when we address subjective assessments of a rational agent about how likely it is that an event will occur.

The sample space, Ω , is the set of all elementary events (e.g. in throwing a die, there are six elementary events), \mathcal{F} is the σ -algebra over Ω . It contains all possible events (e.g. throwing six is a possible event, but so is throwing an even number), and P is a probability function associated with every event, E , in \mathcal{F} . The axioms are

1. The probability of any event is non-negative: $P(E) \geq 0$.
2. The probability of the sample space is one: $P(\Omega) = 1$.
3. The probability of the union of any countable sequence of disjoint events is equal to the sum of the probabilities of all those events:

$$P(E_1 \cup E_2 \cup E_3 \dots) = P(E_1) + P(E_2) + P(E_3) + \dots$$

The limitation to countable sequences is called σ -additivity. From these axioms one can prove that the probability of any event cannot be greater than 1, and that the probability of the null set is 0. Moreover, if E_1 and E_2 are *not* disjoint, one can also prove that

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2). \quad (1)$$

If you know the probabilities of every event in \mathcal{F} , then you know the probability distribution.

Glenn Shafer was interested in situations where one does not know the whole probability distribution, in short where there is some ignorance.¹¹ He introduced the notion of the *mass* of the proposition E , $m(E)$. Bearing in mind that E is a set of possible worlds (see footnote 3), we may interpret $m(E)$ as the weight of evidence supporting only the claim that the actual world belongs to E . Note that this makes room for ignorance in that it leaves open to which particular subset of E the actual world belongs. The mass of the null set is zero, $m(\emptyset) = 0$; and the sum of the masses of all the subsets (and not only the subsets in the σ -algebra) contained in the sample space is one:

$$\sum_{C \subseteq \Omega} m(C) = 1.$$

¹⁰ We thank an anonymous referee for the advice to add this section as a service to readers who are less familiar with the material.

¹¹ Shafer (1976).

The belief in E is defined as the sum of the masses of all the subsets of E :

$$Bel(E) = \sum_{C \subseteq E} m(C).$$

The axioms for Shafer belief functions can now be stated as follows:

1. The belief in any proposition is non-negative: $Bel(E) \geq 0$.
2. The belief in the disjunction of all the propositions (the sample space) is one, $Bel(\Omega) = 1$, and the belief in the impossibility is zero, $Bel(\emptyset) = 0$.
3. The belief in a disjunction is greater than or equal to the sum of the beliefs in the disjuncts, minus the belief in their conjunction:

$$Bel(E_1 \cup E_2) \geq Bel(E_1) + Bel(E_2) - Bel(E_1 \cap E_2). \quad (2)$$

Formula (2) is like (1), except that the equality has been replaced by an inequality.¹² In the special case when there is no ignorance, all the nonzero masses range over singleton sets, and the inequality in (2) becomes an equality. So in that case the belief function reduces to a probability function.

Arthur P. Dempster introduced a rule for combining two sets of mass assignments.¹³ Since this rule has certain undesirable features we shall not use it. Like Meester and Kerkvliet (see their Definition 2.2), we will use instead Fagin and Halpern's conditional belief function:

$$Bel(E_k | E_{k+1}) = \frac{Bel(E_k \cap E_{k+1})}{Bel(E_k \cap E_{k+1}) + 1 - Bel(E_k \cup E_{k+1}^c)}, \quad (3)$$

where E_{k+1}^c is the complement of E_{k+1} in Ω .¹⁴ This somewhat resembles the definition of a conditional probability in the Kolmogorov system, namely

$$P(E_k | E_{k+1}) = \frac{P(E_k \cap E_{k+1})}{P(E_{k+1})}. \quad (4)$$

The numerators in (3) and (4) are similar, except that a belief function occurs in the one and a probability function in the other; but the denominators are very different. Nevertheless, when there is no ignorance, the belief function reduces to a probability

¹² It is possible to generalize (2) so that more than two events are involved, and Shafer did so. However this limited axiom is enough for our purposes, since we will never have to consider unions of more than two events.

¹³ Dempster (1967), Dempster (1968).

¹⁴ Fagin and Halpern (1989).

function, as we mentioned above, and in this case the denominator in (3) is equal to the denominator in (4).¹⁵

In the next two sections we discuss the two examples that have been introduced by Meester and Kerkvliet. Despite the differences in the details, we will show that in both examples the belief and the probability values are equal: $Bel(E_0) = P(E_0) = \frac{1}{2}$.

3 First example

Meester and Kerkvliet introduce their Example 3.1 first abstractly in terms of Shafer mass functions, and then they offer a concrete model. The model features a fair tetrahedral die that is tossed randomly and a lamp that is either on or off at different times $t_k, k = 0, 1, 2, 3, \dots$. Define sets S_j and E_k as follows:

- $S_j =$ ‘the tetrahedron has landed on side j ’, where j can be any one of 1, 2, 3, or 4.
- $E_k =$ ‘the lamp is on at time t_k ’, where k can be any non-negative integer, 0, 1, 2, etc.

The following rules apply. If the tetrahedron has landed on . . .

- . . . side 1, the lamp is on at t_k if k is even, but off if k is odd [1]
- . . . side 2, the lamp is on at t_k if k is odd, but off if k is even [2]
- . . . side 3, the lamp is on at t_k for all k [3]
- . . . side 4, the state of the lamp is unknown. [4]

Meester and Kerkvliet first ask: “To what degree would you believe that the lamp is on at time t_0 ?”, in other words, what is the value of $Bel(E_0)$? Their answer is derived from the Shafer masses

$$\begin{aligned}
 m(E_0 \cap E_1^c \cap E_2 \cap E_3^c \cap \dots) &= \frac{1}{4} & [1] \\
 m(E_0^c \cap E_1 \cap E_2^c \cap E_3 \cap \dots) &= \frac{1}{4} & [2] \\
 m(E_0 \cap E_1 \cap E_2 \cap E_3 \cap \dots) &= \frac{1}{4} & [3] \\
 m(\Omega) &= \frac{1}{4} & [4] \qquad (5)
 \end{aligned}$$

¹⁵ Proof:

$$\begin{aligned}
 P(E_k \cap E_{k+1}) + 1 - P(E_k \cup E_{k+1}^c) &= P(E_k \cap E_{k+1}) + 1 - P(E_k) - P(E_{k+1}^c) + P(E_k \cap E_{k+1}^c) \\
 &= 1 - P(E_{k+1}^c) = P(E_{k+1}).
 \end{aligned}$$

where we recall that Ω is the set of all outcomes, so that the last entry corresponds to the state of ignorance when the tetrahedron lands 4. We find

$$Bel(E_0) = \sum_{C \subseteq E_0} m(C) = \frac{1}{4} + 0 + \frac{1}{4} + 0 = \frac{1}{2}. \tag{6}$$

They next ask “What is your conditional belief on the lamp being on at time t_k , given that it is *on* at time t_{k+1} ?”, and subsequently “What is your conditional belief that it is on at time t_k , given that it is *off* at time t_{k+1} ?”. In other words, what are the values of $Bel(E_k|E_{k+1})$ and $Bel(E_k|E_{k+1}^c)$? We will distinguish conditional belief functions from conditional probability functions by using a and b for the former and α and β for the latter:

$$\begin{aligned} a &= Bel(E_k|E_{k+1}) & b &= Bel(E_k|E_{k+1}^c) \\ \alpha &= P(E_k|E_{k+1}) & \beta &= P(E_k|E_{k+1}^c). \end{aligned}$$

In Example 3.1 the rules for the tetrahedron are symmetric between even and odd values of k . This means that the conditional belief functions and the conditional probability functions are independent of k , which we express by saying that the regress is uniform. (As we will see, this is not so in Meester and Kerkvliet’s Example 3.2 which we discuss in the next section.)

Using formula (3) for the conditional belief in E_k given E_{k+1} , Meester and Kerkvliet calculate the values of a and b in Example 3.1:

$$a = Bel(E_k|E_{k+1}) = \frac{1}{3} \quad b = Bel(E_k|E_{k+1}^c) = \frac{1}{2}. \tag{7}$$

They then write

“Hence in this example, $a = \frac{1}{3}$ and $b = \frac{1}{2}$ so in a classical situation ...the belief in E_0 would be $\frac{3}{7}$.”¹⁶

What they mean is the following. We have shown¹⁷ that for a uniform Kolmogorovian probabilistic regress, $P(E_0)$ equals $\frac{\beta}{1-\alpha+\beta}$. If in the probabilistic case α and β were to take the same values as a and b in the case with belief functions:

$$\alpha = \frac{1}{3} \quad \beta = \frac{1}{2},$$

then it would follow that

$$P(E_0) = \frac{\beta}{1 - \alpha + \beta} = \frac{\frac{1}{2}}{1 - \frac{1}{3} + \frac{1}{2}} = \frac{3}{7}. \tag{8}$$

¹⁶ Meester and Kerkvliet (2019), Sect. 3, Example 3.1. By ‘classical’ they mean ‘Kolmogorov-probabilistic’.

¹⁷ (Peijnenburg 2007; Atkinson and Peijnenburg 2017), Subsection 3.4.

Hence their conclusion: whereas in Example 3.1 the agent’s degree of rational belief in E_0 is $\frac{1}{2}$ when computed with belief functions, with probability functions it would be $\frac{3}{7}$.

This suggests that in Example 3.1 the value of $P(E_0)$ would be $\frac{3}{7}$. However, in Example 3.1 the combination $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$ is *logically impossible*, as we will shortly demonstrate. The result (8) in fact refers to a *different* situation than that which yielded (6), and so to a different situation than the one described in Example 3.1. That is to say, the tetrahedron model by Meester and Kerkvliet, with its specific rules for each of the four options, fits (6), but it does not fit (8). A model that does fit (8) would need seven rather than four options (we will come back to this in Sect. 5). The masses corresponding to (8) are not the same as those in (5), and this means that the situations are not the same.

What would happen if we were to apply a probabilistic analysis to the *same* situation as the one to which we applied a Shafer analysis, namely Example 3.1? As intimated above, we would then run into a contradiction, for in Example 3.1 it cannot be simultaneously true that

$$\alpha = P(E_k|E_{k+1}) = \frac{1}{3} \quad \text{and} \quad \beta = P(E_k|E_{k+1}^c) = \frac{1}{2}.$$

To show this, we first calculate

$$\alpha \equiv P(E_0|E_1) = \frac{P(E_0 \cap E_1)}{P(E_1)}.$$

According to the rules of the tetrahedral model, if S_1 , then the lamp is off at t_1 , so E_1 is false in this case. If S_2 or S_3 , then E_1 is true. If S_4 , we have no information about the state of the lamp; and we will express this ignorance by supposing the probability of the lamp’s being on at t_1 when S_4 to be p , which can have any value in $[0, 1]$. Since the tetrahedron is fair, $P(E_1) = \frac{1}{4}(0 + 1 + 1 + p) = \frac{1}{4}(2 + p)$.

We know that $E_0 \cap E_1$ is false if S_1 or S_2 , but true if S_3 . If S_4 , then the probability of $E_0 \cap E_1$ is qp , where q is the probability that the lamp is on at t_0 , which can also take on any value in $[0, 1]$.¹⁸ Note that we do not need to assume equality of q and p . Since the tetrahedron is fair, $P(E_0 \cap E_1) = \frac{1}{4}(0 + 0 + 1 + qp) = \frac{1}{4}(1 + qp)$. We therefore conclude that

$$\alpha = \frac{1 + qp}{2 + p}. \tag{9}$$

Next we calculate

$$\beta \equiv P(E_0|E_1^c) = \frac{P(E_0 \cap E_1^c)}{P(E_1^c)}.$$

¹⁸ When conditioned on S_4 , E_0 and E_1 are independent:
 $P(E_0 \cap E_1|S_4) = P(E_0|S_4)P(E_1|S_4)$.

Now $P(E_1^c) = 1 - P(E_1) = \frac{1}{4}(2 - p)$, and $P(E_0 \cap E_1^c) = \frac{1}{4}(1 + 0 + 0 + q(1 - p)) = \frac{1}{4}(1 + q(1 - p))$. Hence

$$\beta = \frac{1 + q(1 - p)}{2 - p} . \tag{10}$$

We can now prove that $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$ cannot be simultaneously true. For if $\alpha = \frac{1}{3}$, we see from (9) that

$$2 + p = 3(1 + qp) \quad \text{that is} \quad p = 1 + 3qp .$$

Since p , being a probability value, cannot be greater than one, it follows that q or p must be zero. The option that p is zero is inconsistent with the latter equation, so q must vanish, which means that $p = 1$. With the values $q = 0$ and $p = 1$, we see from (10) that β must be equal to one. Ergo $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$ is impossible.

In Example 3.1 it is in fact the case that $P(E_0) = Bel(E_0) = \frac{1}{2}$. This can be explained as follows. The agent in Example 3.1 is completely rational. For the Shafer functions described by Meester and Kerkvliet this means not only that the agent bases her beliefs about the lamp and the tetrahedron on relative frequencies, but also that she settles on what she knows for sure. Referring to their formula for conditional belief, see Eq.(3) above, Meester and Kerkvliet write:

The expression [(3)] has a very intuitive rationale, as follows. If we were to repeat the experiment many times, then we can consider the frequency occurring in the subsequence of outcomes that $[E_{k+1}]$ occurs. The expression in [(3)] turns out to be *the minimum such relative frequency* of $[E_k]$ that we can be sure of, given the information contained in the basic belief assignment.¹⁹

We follow Meester and Kerkvliet in adopting the minimum relative frequency that we can be sure of. This amounts to setting $q = 0$ and $p = 0$ in (9) and (10), which leads to $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$.²⁰ The standard formula for a uniform regress then yields

$$P(E_0) = \frac{\beta}{1 - \alpha + \beta} = \frac{\frac{1}{2}}{1 - \frac{1}{2} + \frac{1}{2}} = \frac{1}{2} . \tag{11}$$

¹⁹ Meester and Kerkvliet (2019), Sect. 2, our italics. Cf. their dialogue between T. and R., which makes clear that when T. says “I have no confidence at all for the lamp being on corresponding to rolling 4”, he means that his degree of belief in the lamp being on in the case of S_4 is the minimum value, i.e. zero.

²⁰ The formula (9) was worked out for $P(E_0|E_1)$. Although $P(E_1|E_2)$, $P(E_2|E_3)$, and so on, will be given by the same formula, they may have different values for p and q . However, once we settle for the minimum relative frequency we can be sure of, all values of p and q equal zero. Similar considerations apply to $P(E_0|E_1^c)$, $P(E_1|E_2^c)$, $P(E_2|E_3^c)$, and so on. See also footnote 24.

So Example 3.1 yields the same numerical value for the probability of E_0 as for the Shafer belief in E_0 , namely one half.²¹

4 Second example

As they did with Example 3.1, Meester and Kerkvliet first introduce Example 3.2 purely abstractly in terms of Shafer masses. But while in Example 3.1 the abstract system is followed by an interpretation in terms of a concrete model, viz. with a tetrahedron, this is not so for Example 3.2. However, we can construct a model by tweaking the tetrahedron case a little.

Imagine again a fair tetrahedron that is tossed randomly, and a lamp that is on or off. This time the rules are a bit different from those in Example 3.1. If the tetrahedron has landed on ...

- ... side 1, the lamp is on at t_k if k is even, but off if k is odd [1]
- ... side 2, the lamp is on at t_k if k is odd, but off if k is even [2]
- ... side 3, the lamp is off at t_k for all k [3]
- ... side 4, the lamp is on at t_k if k is even, but if k is odd, the state of the lamp is unknown. [4]

The rules for sides 1 or 2 are the same as they were in Example 3.1; but for side 3 the rule is opposite to what it was before; and, more importantly, for side 4 the even/odd symmetry has been broken, because ignorance now only reigns when k is odd. The regress of Example 3.2 is therefore non-uniform, which will turn out to be relevant when we analyze the example probabilistically.

The Shafer masses are

$$\begin{aligned}
 m(E_0 \cap E_1^c \cap E_2 \cap E_3^c \cap \dots) &= \frac{1}{4} & [1] \\
 m(E_0^c \cap E_1 \cap E_2^c \cap E_3 \cap \dots) &= \frac{1}{4} & [2] \\
 m(E_0^c \cap E_1^c \cap E_2^c \cap E_3^c \cap \dots) &= \frac{1}{4} & [3] \\
 m(E_0 \cap E_2 \cap E_4 \cap E_6 \cap \dots) &= \frac{1}{4} & [4] \tag{12}
 \end{aligned}$$

where in the last entry all the k in E_k are even, since if k is odd, the state of the lamp is unknown.

²¹ We have for the sake of argument ignored the fact that Meester and Kerkvliet’s Example 3.1 is actually not a regress. For when $\alpha = \beta$,

$$P(E_0) = \alpha P(E_1) + \beta P(E_1^c) = \beta (P(E_1) + P(E_1^c)) = \beta.$$

This means that E_1 , and all the higher E_k , drop out, so we cannot speak of a regress. As we will see in the next section, their Example 3.2 does not suffer from this defect. In that example α and β have different values, and there is a real regress.

A rational agent will calculate as the belief functions:

$$a = Bel(E_k|E_{k+1}) = 0 \quad b = Bel(E_k|E_{k+1}^c) = \frac{1}{2},$$

for all k . The unconditional belief in E_0 is: $Bel(E_0) = \frac{1}{4} + 0 + 0 + \frac{1}{4} = \frac{1}{2}$.

As in Example 3.1, Meester and Kerkvliet set the values of the conditional probabilities equal to the values of the conditional beliefs:

$$\alpha = P(E_k|E_{k+1}) = 0 \quad \beta = P(E_k|E_{k+1}^c) = \frac{1}{2}.$$

These values of α and β would lead to

$$P(E_0) = \frac{\beta}{1 - \alpha + \beta} = \frac{\frac{1}{2}}{1 - 0 + \frac{1}{2}} = \frac{1}{3}. \tag{13}$$

They conclude again that $Bel(E_0)$ and $P(E_0)$ have different values. In their own words:

“Hence in this example we have $a = 0$ and $b = \frac{1}{2}$. In the classical case this would lead to probability $\frac{1}{3}$...that E_0 is true, but we have obtained $Bel(E_0) = \frac{1}{2}$.”²²

However, as in the previous section, these two cases refer to different situations. When Meester and Kerkvliet conclude that $Bel(E_0) = \frac{1}{2}$ and $P(E_0) = \frac{1}{3}$, this is only true in the sense that $Bel(E_0) = \frac{1}{2}$ in Example 3.2 and that $P(E_0) = \frac{1}{3}$ in *some other* example, not further specified, where incidentally the conditional probability functions numerically agree with the conditional belief functions.

Let us see what happens if we perform a probabilistic analysis on the same Example 3.2. If S_1 or S_3 , then the lamp is off at t_1 , so E_1 is false in those cases, but if S_2 , then E_1 is true. If S_4 , we have no information about the state of the lamp at t_1 , hence we set the probability of E_1 given S_4 equal to some p in $[0,1]$. Since the tetrahedron is fair, $P(E_1) = \frac{1}{4}(0 + 1 + 0 + p) = \frac{1}{4}(1 + p)$. $E_0 \cap E_1$ is false if S_1 or S_2 or S_3 . If S_4 , then we do not know whether $E_0 \cap E_1$ is false or true; we do know, however, that E_0 is true and that there is a certain probability p that E_1 is true as well. Hence: $P(E_0 \cap E_1) = \frac{1}{4}p$. Since the conditional probabilities differ for even and odd values of k , the symmetry breaking in Example 3.2 is now relevant, and we therefore add an index to α and β . We conclude that

$$\alpha_0 \equiv P(E_0|E_1) = \frac{\frac{1}{4}p}{\frac{1}{4}(1 + p)} = \frac{p}{1 + p}. \tag{14}$$

We have $P(E_1^c) = 1 - P(E_1) = \frac{1}{4}(3 - p)$. Note that $E_0 \cap E_1^c$ is true if S_1 , but false if S_2 or S_3 , whereas if S_4 its probability is $1 - p$, so $P(E_0 \cap E_1^c) = \frac{1}{4}(1 + 1 - p) = \frac{1}{4}(2 - p)$.

²² Meester and Kerkvliet (2019), Sect. 3, Example 3.2. Again, ‘classical’ here means ‘Kolmogorov-probabilistic’.

Hence

$$\beta_0 \equiv P(E_0|E_1^c) = \frac{2 - p}{3 - p}. \tag{15}$$

Meester and Kerkvliet’s values for α and β in Example 3.2 are again inconsistent, for if $\alpha_0 = 0$ we see from (14) that $p = 0$. But then from (15) it follows that $\beta_0 = \frac{2}{3}$, not $\frac{1}{2}$, as Meester and Kerkvliet would have it. Similar formulae apply to α_k and β_k for all the even values of k ; but for the odd values we need to perform a separate calculation.

To calculate α_1 , we note first that E_2 is true if S_1 or S_4 , false if S_2 or S_3 , so $P(E_2) = \frac{1}{4}(1 + 0 + 0 + 1) = \frac{1}{2}$. On the other hand, $E_1 \cap E_2$ is false if S_1, S_2 or S_3 , and true with probability p if S_4 ,²³ so

$$\alpha_1 \equiv P(E_1|E_2) = \frac{\frac{1}{4}P}{\frac{1}{2}} = \frac{1}{2}P. \tag{16}$$

We have $P(E_2^c) = 1 - P(E_2) = \frac{1}{2}$. Moreover, $E_1 \cap E_2^c$ is true if S_2 , false if S_1, S_3 or S_4 , thus

$$\beta_1 \equiv P(E_1|E_2^c) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}. \tag{17}$$

Similar formulae apply to α_k and β_k for all the odd values of k . Evidently the values of α_k and β_k are different for even or odd k — a significant contrast with the Shafer belief functions, which are the same for even and odd k .

Following Meester and Kerkvliet again in settling on the minimum relative frequency that we can be sure of, we set $p = 0$ in (14), (15), (16) and (17), resulting in $\alpha_k = 0$ and $\beta_k = \frac{2}{3}$ for even k , and $\alpha_k = 0$ and $\beta_k = \frac{1}{2}$ for odd k .²⁴

Since β_k is different for even and odd k , the regress is not uniform. So in contrast to what was the case in Example 3.1, the formula $\frac{\beta}{1-\alpha+\beta}$ cannot immediately be used to evaluate $P(E_0)$. To circumvent this difficulty, we combine the equations for the even and odd k in such a way that $P(E_k)$ is eliminated for all odd k , and we retain only the even values. This composite regress is uniform, so the above formula *can* now be used.

The rule of total probability for E_0 is

$$\begin{aligned} P(E_0) &= \alpha_0 P(E_1) + \beta_0 P(E_1^c) \\ &= \alpha_0 P(E_1) + \beta_0 [1 - P(E_1)] \\ &= \beta_0 + \gamma_0 P(E_1), \end{aligned}$$

²³ This is the same probability as the one in (14) and (15), since it is also the chance that the lamp is on at t_1 if S_4 .

²⁴ The formula (14) applies to α_0 and (16) applies to α_1 . Similarly, α_2, α_4 , and so on, are given by (14), and α_3, α_5 , and so on, are given by (16), with possibly different values for p . Again we settle on the minimum relative frequency we can be sure of, so all values of p equal zero. Similar considerations pertain to $\beta_2, \beta_3, \beta_4$, and so on. See also footnote 20.

where $\gamma_0 = \alpha_0 - \beta_0$. We now eliminate $P(E_1)$ by using the rule of total probability for $P(E_1)$, which can be written in the form

$$P(E_1) = \beta_1 + \gamma_1 P(E_2),$$

where $\gamma_1 = \alpha_1 - \beta_1$. The result is

$$P(E_0) = \beta_0 + \gamma_0 \beta_1 + \gamma_0 \gamma_1 P(E_2), \tag{18}$$

which is equivalent to

$$P(E_0) = \hat{\alpha} P(E_2) + \hat{\beta} P(E_2^c), \tag{19}$$

where $\hat{\alpha} = \beta_0 + \gamma_0 \alpha_1 = \frac{2}{3}$ and $\hat{\beta} = \beta_0 + \gamma_0 \beta_1 = \frac{1}{3}$.²⁵ All the $P(E_k)$ for even k obey a rule of total probability like (19): they constitute a regress in which the odd k have been eliminated, and where the effective conditional probabilities are $\hat{\alpha}$ and $\hat{\beta}$. We can now use the standard formula:

$$P(E_0) = \frac{\hat{\beta}}{1 - \hat{\alpha} + \hat{\beta}} = \frac{\frac{1}{3}}{1 - \frac{2}{3} + \frac{1}{3}} = \frac{1}{2}.$$

This agrees with the result of direct calculation.²⁶

5 Summary and conclusion

In a recent paper, Ronald Meester and Timber Kerkvliet have given two examples of a probabilistic regress in terms of Shafer belief functions. In each of the examples, the regress incorporates a well-defined nonzero number for the unconditional belief in the target proposition, $Bel(E_0)$. On the basis of these examples, Meester and Kerkvliet claim that the value of $Bel(E_0)$ may differ from that of $P(E_0)$, the unconditional probability of the target. In the present paper we explained that this is misleading. A detailed analysis with probability functions shows that in both examples the values of $P(E_0)$ and $Bel(E_0)$ are the same. The only way in which these values can differ from one another is if they refer to different examples, but of course that makes the claim unsurprising.

The first example of Meester and Kerkvliet, Example 3.1, involves a fair tetrahedron with specific rules pertaining to each of the four possible outcomes. In this example, $a = \frac{1}{3}$ and $b = \frac{1}{2}$, where a and b are conditional belief functions, and the rational

²⁵ To get from (18) to (19), one can write $P(E_0) = (\beta_0 + \gamma_0 \beta_1) [P(E_2) + P(E_2^c)] + \gamma_0 \gamma_1 P(E_2)$; it then suffices to gather the two pieces involving $P(E_2)$ together.

²⁶ Meester and Kerkvliet’s Theorem 3.3(b) shows that there exists an infinite number of examples like 3.1 and 3.2. However, each of these examples is open to our criticism. None of them show that a regress can have different solutions in a particular situation, dependent on whether one does a Kolmogorov or a Shafer analysis. All the examples again only show that a regress may have different solutions in different situations.

agent concludes that his degree of belief in E_0 is one half: $Bel(E_0) = \frac{1}{2}$. Meester and Kerkvliet argue that with $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$, where α and β are conditional probability functions, the value would be different: $P(E_0) = \frac{3}{7}$. As we have shown, however, they thereby change the example, for $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$ cannot be simultaneously true in Example 3.1.

Here is an example in which $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$ are simultaneously true. Consider a fair roulette wheel with numbers 1 to 7, the probability that the wheel stops at a given number being $\frac{1}{7}$. In analogy with the tetrahedron model, define sets S_j and E_k as follows:

S_j = ‘the roulette wheel has stopped at number j ’, where j can be any one of 1, 2, ..., 7.

E_k = ‘the lamp is on at time t_k ’, where k can be any non-negative integer, 0, 1, 2, etc.

The following rules apply. If the roulette wheel has stopped at ...

- ...numbers 1 or 2, the lamp is on at t_k if k is even, but off if k is odd
- ...numbers 3 or 4, the lamp is on at t_k if k is odd, but off if k is even
- ...number 5, the lamp is on at t_k for all k
- ...numbers 6 or 7, the lamp is off at t_k for all k .

With these rules one can check that $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{2}$ and $P(E_0) = \frac{3}{7}$.

This roulette example is quite different from Example 3.1. Probability functions are special cases of belief functions, so like the latter they can be described in terms of mass functions. The point is that the mass functions of which our roulette example serves as a model differ from the mass functions in the tetrahedron model of Example 3.1.

Similar considerations apply to Meester and Kerkvliet’s second example, Example 3.2. In this case it turns out that one can make do with the fair tetrahedron, but with rules that are other than those given by Meester and Kerkvliet. Namely, if the tetrahedron has landed on ...

- ...sides 1 or 2, the lamp is on at t_k if k is even, but off if k is odd
- ...side 3, the lamp is on at t_k if k is odd, but off if k is even
- ...side 4, the lamp is off at t_k for all k .

One can check that $P(E_0) = \frac{1}{2}$, and $\alpha_n = 0$ for all n , while $\beta_n = \frac{2}{3}$ for even n , but $\beta_n = \frac{1}{2}$ for odd n . The masses are not the same as those in Example 3.2, so this configuration corresponds once more to a different situation.

In sum, given the masses, both in Example 3.1 and Example 3.2 the value of the unconditional probability of the target, $P(E_0)$, is identical to the value of the unconditional belief, $Bel(E_0)$.

That this identity is a general one, and not one that just applies to these two examples, can be readily seen as follows. By adding up the relevant masses, one can calculate directly both $Bel(E_0)$ and $P(E_0)$. It will always be the case that $Bel(E_0)$ is a number between 0 and 1; but $P(E_0)$ will in general be a function of variable probabilities p_n that can take on any value in the unit interval. However, with Meester and Kerkvliet’s definition of the conditional belief functions, we have seen that the minimum relative

frequencies are effectively chosen: “The expression in [(3)] turns out to be the minimum such relative frequency of E that we can be sure of”. This means that all the p_n are zero, and in that case $P(E_0) = Bel(E_0)$.

Acknowledgements We thank Ronald Meester and Timber Kerkvliet for very helpful and clarifying discussions. We also thank two anonymous referees whose suggestions led to considerable improvement.

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

- Aikin, S. F. (2011). *Epistemology and the regress problem*. New York, Oxford: Routledge.
- Aikin, S. F., & Peijnenburg, J. (2014). The regress problem. *Metaphilosophy*, 45, 139–145.
- Atkinson, D., & Peijnenburg, J. (2017). Fading foundations. *Probability and the regress problem*. Dordrecht: Springer. <https://doi.org/10.1007/978-3-319-58295-5>.
- Bonjour, L. (1985). *The structure of empirical knowledge*. Cambridge: Harvard University Press.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38, 325–339.
- Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society Series B*, 30, 205–247.
- Fagin, R., & Halpern, J. Y. (1989). Uncertainty, belief and probability. In N. Y. Sridharan (Ed.), *Proceedings of the eleventh international joint conference on artificial intelligence* (Vol. 1, pp. 1161–1167). San Mateo: Morgan Kaufmann Publishers.
- Herzberg, F. (2010). The consistency of probabilistic regresses. *Studia Logica*, 94, 331–345.
- Huemer, M. (2016). *Approaching infinity*. Basingstoke, New York: Palgrave MacMillan.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.
- Lewis, C. I. (1929). *Mind and the world-order. An outline of a theory of knowledge*. New York: C. Scribner’s Sons.
- Meester, R., & Kerkvliet, T. (2019). The infinite epistemic regress problem has no unique solution. *Synthese*. <https://doi.org/10.1007/s11229-019-02383-7>.
- Pearl, J. (1988). On probability intervals. *International Journal of Approximate Reasoning*, 2, 211–216.
- Peijnenburg, J. (2007). Infnitism regained. *Mind*, 116, 597–602.
- Rescher, N. (2010). *Infinite regress. The theory and history of varieties of change*. New Brunswick: Transaction Publishers.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.
- Shafer, G. (1981). Constructive probability. *Synthese*, 48, 1–60.
- Turri, J., & Klein, P. D. (Eds.). (2014). *Ad infinitum. New essays on epistemological infinitism*. Oxford: Oxford University Press.

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