

# *For Bayesians, Rational Modesty Requires Imprecision*

Brian Weatherson

2016

*Abstract:* Gordon Belot has recently developed a novel argument against Bayesianism. He shows that there is an interesting class of problems that, intuitively, no rational belief forming method is likely to get right. But a Bayesian agent's credence, before the problem starts, that she will get the problem right has to be 1. This is an implausible kind of immodesty on the part of Bayesians. My aim is to show that while this is a good argument against traditional, precise Bayesians, the argument doesn't neatly extend to imprecise Bayesians. As such, Belot's argument is a reason to prefer imprecise Bayesianism to precise Bayesianism.

GORDON Belot (2013) has recently developed a novel argument against Bayesianism. He shows that there is an interesting class of problems that, intuitively, no rational belief forming method is likely to get right. But a Bayesian agent's credence, before the problem starts, that she will get the problem right has to be 1. This is an implausible kind of *immodesty* on the part of Bayesians.<sup>1</sup> My aim is to show that while this is a good argument against traditional, precise Bayesians, the argument doesn't neatly extend to imprecise Bayesians. As such, Belot's argument is a reason to prefer imprecise Bayesianism to precise Bayesianism.

For present purposes, the precise Bayesian agent has just two defining characteristics. First, their credences in all propositions are given by a particular countably additive probability function. Second, those credences are updated by conditionalisation as new information comes in. These commitments are quite strong in some respects. They say that there is a single probability function that supplies the agent's credences no matter which question is being investigated, and no matter how little evidence the agent has before the investigation is started. The everyday statistician, even one who is sympathetic to Bayesian approaches, may feel no need to sign up for anything this

<sup>1</sup>There is another sense of immodesty that is often discussed in the literature, going back to Lewis (1971). This is the idea that some agents think their attitudes are optimal by some standards; these are the immodest ones. And often, it is held that not being self-endorsing in this way is a coherence failure Elga (2010). I don't think this kind of immodesty is rationally required, for reasons set out by Miriam Schoenfield (2015) and Maria Lasonen-Aarnio (2015), but in any case that's not the kind of modesty that's at issue in Belot's argument.

strong. But many philosophers seem to be interested in varieties of Bayesianism that are just this strong. For instance, there has been extensive discussion in recent epistemology of whether various epistemological approaches, such as dogmatism, can be modeled within the Bayesian framework, with the background assumption being that it counts against those approaches if they cannot.<sup>2</sup> In these debates, the issue is not whether the Bayesian approach works in the context of a well-defined question and a substantial evidential background, but whether it does so for all questions in all contexts. Indeed, the assumption is that it does, and epistemological theories inconsistent with it are false. So the precise Bayesian is a figure of some interest, at least in epistemology.

The imprecise Bayesian doesn't have a single probability function for their credences. Rather, they have a representor consisting of a set of probability functions. The agent is more confident in  $p$  than  $q$  just in case  $\Pr(p) > \Pr(q)$  for every  $\Pr$  in this representor.<sup>3</sup> Just like the precise Bayesian, the imprecise Bayesian updates by conditionalisation; their new representor after an update is the result of conditionalising every member of the old representor with the new information. The added flexibility in imprecise Bayesianism will allow us to develop a suitably modest response to Belot's puzzle.

## 1 The Puzzle

The set up Belot uses is this. An agent,  $A$ , will receive a data stream of 0s and 1s. The data stream will go on indefinitely. I will use  $\mathbf{x}$  for the (infinite) sequence of data she would (eventually) get,  $x_k$  for the  $k$ th element of this sequence, and  $\mathbf{x}_k$  for the sequence consisting of the first  $k$  elements of the stream. These variables are, as usual, rigid designators. I'll also use the capitalised  $\mathbf{X}$  and  $\mathbf{X}_k$  as random variables for the sequence itself, and for the first  $k$  elements of the sequence, respectively. So  $\mathbf{X} = \mathbf{x}$  is the substantive and true claim that the sequence that will be received is actually  $\mathbf{x}$ . And  $\mathbf{X}_k = \mathbf{x}_k$  is the substantive and true claim that the first  $k$  elements of that sequence are  $\mathbf{x}_k$ . Propositions of this form will play a major role below, since they summarise the evidence the agent has after  $k$  elements have been revealed. I'll use  $+$  as a sequence concatenation operator, so  $\mathbf{y} + \mathbf{z}$  is the sequence consisting of all of  $\mathbf{y}$ , followed by all of  $\mathbf{z}$ .

<sup>2</sup>For dogmatism, see Pryor (2000). The canonical argument that it is inconsistent with Bayesianism is White (2006).

<sup>3</sup>Note that this formulation leaves it open which side of the biconditional is explanatorily prior. I'm going to defend a view on which the left hand side, i.e., the comparative confidences, are more explanatorily basic than the facts about what is in the agent's representor. I say a little more about why I take this stand in footnote 7. For much more detail on varieties of imprecise Bayesianism, see Walley (1991), from whom I take the view that the representor and its members are much less explanatorily important than the comparative judgments the agent makes.

Belot is interested in a quite general puzzle, but I'll focus for most of the paper on a very specific instance of the puzzle. (We'll return to the more general puzzle in the last section.) We're going to look at the agent's evolving credence that  $\mathbf{X}$  is *periodic*. Let  $p$  be the proposition that  $\mathbf{X}$  is periodic, since we'll be returning to that proposition a lot. And let's start by assuming the agent is a precise Bayesian, to see the challenge Belot develops.

Say that the agent **succeeds** just in case her credence in  $p$  eventually gets on the correct side of  $\frac{1}{2}$ , and stays there. (The correct side is obviously the high side if  $p$  is true, and the low side otherwise.) That is, if  $v$  is the truth value function, it succeeds just in case this is true.<sup>4</sup>

$$\exists n \forall m \geq n: |v(p) - Cr(p \mid \mathbf{X}_m = \mathbf{x}_m)| < \frac{1}{2}$$

The agent **fails** otherwise. Given the assumption that the agent is a classical Bayesian, we can step back from evaluating the agent and evaluate her prior probability function directly. So a prior  $\text{Pr}$  succeeds relative to  $\mathbf{x}$  just in case this is true.

$$\exists n \forall m \geq n: |v(p) - \text{Pr}(p \mid \mathbf{X}_m = \mathbf{x}_m)| < \frac{1}{2}$$

This is reasonably intuitive; the agent is going to get a lot of data about  $\mathbf{X}$ , and it is interesting to ask whether that data eventually lets her credence in  $p$  get to the right side of  $\frac{1}{2}$ .

Given these notions of success and failure, we can naturally define the success set of a prior (or agent) as the set of sequences it succeeds on, and the failure set as the set of sequences it fails on.

Abusing notation a little, say that  $\mathbf{x}_i \supset \mathbf{x}_k$  iff  $\mathbf{x}_i$  is a sequence that has  $\mathbf{x}_k$  as its first  $k$  entries. Then we can state the first of Belot's conditions on a good Bayesian agent/prior. A prior is **open-minded** just in case this condition holds:

$$\forall \mathbf{x}_k \exists \mathbf{x}_i \supset \mathbf{x}_k, \mathbf{x}_j \supset \mathbf{x}_k: \text{Pr}(p \mid \mathbf{X}_i = \mathbf{x}_i) < \frac{1}{2} \wedge \text{Pr}(p \mid \mathbf{X}_j = \mathbf{x}_j) > \frac{1}{2}$$

That is, no matter what happens, it is possible that the probability of  $p$  will fall below, and possible it will rise above. To motivate the first, consider any situation where the sequence to date has looked periodic. (If it had not looked periodic to date, presumably the probability of  $p$  should already be low.) Now extend that sequence with a large of random noise. At the end of this, it should no longer be probable that the sequence is periodic. On the other hand, assume the sequence has not looked periodic to date. Extend it by repeating  $\mathbf{x}_k$  more than  $k$  times. At the end of this, it should look probable

<sup>4</sup>Belot lets an agent succeed if  $\mathbf{X}$  is periodic, and the credence in  $p$  never drops below  $\frac{1}{2}$ , but I think it's neater to say that the agent is undecided in this case.

that the sequence is periodic (at least for large enough  $k$ ). So open-mindedness looks like a good condition to impose.

The second condition we might impose, though not one Belot names, is *modesty*. Any function might fail. One natural way it might fail is that it might get, to use a term Belot does use, *flummoxed*. It could change its mind infinitely often about whether the sequence is periodic. By definition, open-mindedness entails the possibility of being flummoxed. Given the definitions of success and failure, Pr will fail relative to any  $x$  that flummoxes it. So success is not a priori guaranteed. Now for any function we can work out the set of sequences relative to which it fails. It turns out this will be a rather large set. Indeed, the set of sequences on which any open-minded function succeeds is meagre.<sup>5</sup> Say a function is **modest** if the initial probability it gives to  $\mathbf{X}$  being in its success set is less than 1. Given how large the failure set is, modesty also seems like a good requirement.<sup>6</sup>

The argument for modesty is not that it is an immediate consequence of regularity. It does follow from regularity, but in the case we're considering, regularity is quite implausible. Some sets, even some quite large sets in some sense, will have to be given probability 0. The surprising thing is that a residual set (i.e., the complement of a meagre set) gets probability 0.

It might be thought that modesty here is problematic for the same reason that epistemic modesty is often problematic: it validates Moore-paradoxical thoughts. It's bad to say  $p$ , *but there is a probability that not  $p$* . It's even bad, though as Briggs (2009) points out, not quite bad for the same reasons, to say *Whether I believe  $p$  is true or false tomorrow, there will be a probability I'm false*. Perhaps modesty is a requirement that someone say something like that, and hence is an improper requirement.

But in fact the requirement of modesty is disanalogous to the 'requirement', suggested in the previous paragraph, that agents endorse Moore-paradoxical principles. There isn't anything wrong with saying *Whichever side of one half my credence in  $p$  is tomorrow, there is a probability that the truth will be the other side of one half*. That's not Moore-paradoxical. Indeed, unless one is sure that one's credence in  $p$  tomorrow will be 0 or 1, it is something one should endorse.

Or consider a different example. There will be a sequence of 0s and 1s, but this time there will only be three elements, and the agent will only be shown the first of them tomorrow. Let  $q$  be the proposition that there are more 1s than 0s in the three-element sequence. Say the agent **succeeds** iff tomorrow, after seeing just one element, her credence in  $q$  is the same side of one-half as the truth. And say the agent is **modest** iff, right

<sup>5</sup>A meagre subset of a space is any set built up as a countable union of nowhere dense sets.

<sup>6</sup>Belot goes into much more detail about why modesty is a good requirement to put on a rational prior, but I'm omitting those details since I have very little to add to what Belot says.

now, her credence that she succeeds tomorrow is less than one. There is nothing incoherent about being modest. If her credal distribution today is completely flat, giving  $1/8$  credence to each of the eight possible sequences, she will be modest, for example.

Now this case is somewhat different to the one Belot started with in a couple of respects. On the one hand, we're asking about modesty at a particular point, i.e., tomorrow, rather than over a long sequence. On the other hand, we're asking about whether the agent's credences will be on the right side of one-half after having seen one-third of the data, rather than, as in the original case, after seeing measure zero of the sequence. The first difference makes it easier to be modest, the second difference makes it harder. So the cases are not perfect analogies, but they are similar enough in respect of modesty to make it plausible that if modesty is coherent in this case, as we've shown it is, then it should be coherent in Belot's case as well.

So that's the argument that open-mindedness and modesty are good conditions for priors to satisfy. Here's the worrying result that Belot proves. There are no open-minded modest priors. If  $A$  is a classical Bayesian, she will either have to be closed minded or immodest. Neither seems rational, so it seems that being a classical Bayesian is incompatible with being rational. That is, we can't be precise Bayesians if we accept the following two constraints.

- *Open-Mindedness*: For any initial sequence, there is a continuation after which it seems probable that  $\mathbf{X}$  is periodic, and a continuation after which it seems probable that  $\mathbf{X}$  is not periodic.
- *Modesty*: The initial probability that the agent will succeed, i.e., that their credence in  $p$  will eventually get to the right side of  $1/2$  and stay there, is less than 1.

Since both open-mindedness and modesty are very plausible constraints, it follows that there is no good way to be a precise Bayesian in the face of this puzzle.

## 2 Making the Puzzle Less Precise

What happens, though, if the agent is an imprecise Bayesian? Is there a parallel version of Belot's argument that shows this kind of imprecise Bayesian is necessarily irrational? I'm going to argue that the answer is no.

The first thing we have to do is work out how to redefine the key terms in Belot's argument once we drop the assumption that the agent is a classical Bayesian. There are several ways of formulating our definitions which are equivalent given that assumption, but not equivalent given that the agent is an imprecise Bayesian. There are three major choice points here.

1. What is success?
2. What is open-mindedness?
3. What is modesty?

Assume our agent's credal state is represented by set  $S$  of probability functions. Then there are two natural ways to think about success.

1.  $\forall \text{Pr} \in S: \exists n \forall m \geq n: |v(p) - \text{Pr}(p \mid \mathbf{X}_m = \mathbf{x}_m)| < 1/2$
2.  $\exists n \forall \text{Pr} \in S: \forall m \geq n: |v(p) - \text{Pr}(p \mid \mathbf{X}_m = \mathbf{x}_m)| < 1/2$

The second is obviously stronger than the first, since it involves moving an existential quantifier out in front of a universal quantifier. And there are some natural cases where an agent could succeed on the first definition, and fail on the second. Here's one such case.

Let  $\text{Pr}_0$  be the *fair-coin measure*. According to the fair coin measure, if  $\mathbf{y}$  is any  $k$  length sequence of 0s and 1s we have  $\text{Pr}_0(\mathbf{x}_k = \mathbf{y}) = 2^{-k}$ . Intuitively, it thinks the 0s and 1s are generated by flips of a fair coin, and it won't change its mind about that no matter what happens.

Say a probability function  $\text{Pr}$  is **regular periodic** iff it satisfies these two conditions.

- $\text{Pr}(p) = 1$ .
- For any periodic sequence  $\mathbf{y}$ ,  $\text{Pr}(\mathbf{X} = \mathbf{y}) > 0$ .

Intuitively, these functions are certain that  $X$  is periodic, and assign positive probability to each possible periodic sequence. Now consider the family of functions we get by taking equal weighted mixtures of  $\text{Pr}_0$  with each regular periodic function. Let that family represent the agent's credence. And assume for now that  $\mathbf{X}$  is the sequence  $\langle 0, 0, \dots \rangle$ . Does the agent succeed?

Well, each  $\text{Pr}$  in her representor succeeds. To prove this, it will be helpful to prove a lemma that we'll again have use for below. For this lemma, let  $\text{Pr}_0$  be the fair-coin measure (as already noted),  $\text{Pr}_1$  be any measure such that  $\text{Pr}_1(p) = 1$ , and  $\text{Pr}_2$  be the equal mixture of  $\text{Pr}_0$  and  $\text{Pr}_1$ .

**Lemma 1:**  $\text{Pr}_2(p \mid \mathbf{X}_k = \mathbf{y}_k) > 1/2$  iff  $\text{Pr}_1(\mathbf{X}_k = \mathbf{y}_k) > \text{Pr}_0(\mathbf{X}_k = \mathbf{y}_k)$ .

*Proof.* Let  $\text{Pr}_i(\mathbf{X}_k = \mathbf{y}_k) = a_i$  for  $i \in \{0, 1\}$ . Recall that  $\text{Pr}_0(p) = 0$  and  $\text{Pr}_1(p) = 1$ . Then we can quickly get that  $\text{Pr}_2(p \mid \mathbf{X}_k = \mathbf{y}_k) = a_1 / (a_0 + a_1)$ , from which the lemma immediately follows.  $\square$

For any  $\text{Pr}$  in the agent's representor, there is some  $k$  such that  $\text{Pr}(\mathbf{X} = \langle 0, 0, 0, \dots \rangle) > 2^{-k}$ . So after at most  $k$  0s have appeared,  $\text{Pr}(p)$  will be above  $1/2$ , and it isn't coming back. That means it succeeds. And since  $\text{Pr}$  was arbitrary, it follows that all  $\text{Pr}$  succeed.

But the agent in a good sense doesn't succeed. No matter how much data she gets, there will be  $\Pr$  in her representor according to which  $\Pr(p) < 1/2$ . After all, for any  $k$ , there are regular periodic  $\Pr$  such that the probability of  $x_k$  being  $k$  0s is below  $2^{-k}$ . So if we mix that function with  $\Pr_0$ , we get a function where the most probable continuations of this initial sequence are the random sequences provided by the fair coin measure.

In terms of our definitions of success above, the agent satisfies the first, but not the second. Every function in her representor eventually has the probability of  $p$  go above  $1/2$ . But at any time, there are functions in her representor according to which the probability of  $p$  is arbitrarily low.

Here I think we have to make a distinction between different ways of understanding the formalism of imprecise probabilities. (What follows is indebted to Bradley (2014), especially his section 3.1, but I'm disagreeing somewhat with his conclusions, and following more closely the conclusions of Joyce (2010) and Schoenfield (2012).)

One way of thinking about imprecise credences is that each probability function in the representor is something like an advisor, and the agent who is imprecise simply hasn't settled on which advisor to trust. Call this the **pluralist** interpretation of the formalism. On this interpretation, it is natural to think that what is true of every function is true of the agent.

Another way is to think of the agent's mind as constituted by, but distinct from, the representors. An analogy to keep in mind here is the way that a parliament is constituted by, but distinct from, its members. Keeping with this analogy, call this the **corporate** interpretation of the formalism. Note that corporate bodies will typically have their own rules for how the views of the members will be translated into being views of the whole. Even if every member of the parliament believes that the national cricket team will win its upcoming game, it doesn't follow that the parliament believes that; the parliament only believes what it resolves it has believed.

Now I only want to defend the imprecise Bayesian model on the corporate interpretation.<sup>7</sup> The pluralist interpretation, it seems to me, faces grave difficulties. For one thing, it has a hard time explaining what's wrong with the existential claim "There is a precise number  $x$  such that  $x$  is the probability of  $p$ ". Every advisor believes that, so on the pluralist model the agent does too. (Compare the criticisms of "fanatical supervaluationism" in Lewis (1993).) More relevant to the discussion here, I am following Belot

<sup>7</sup>I have an independent metaphysical reason for preferring the corporate interpretation. I think that comparative confidences, things like being at least as confident in  $p$  as in  $q$ , are metaphysically prior to numerical credences, or even sets of numerical credences. On such a metaphysics, what it is for  $\Pr$  to be in the representor just is for every  $p, q, r, s$ , if the agent is at least as confident in  $p$  given  $q$  as in  $r$  given  $s$ , then  $\Pr(p | q) \geq \Pr(r | s)$ . And it seems, though I won't defend this claim here, that the corporate interpretation fits more naturally with the idea that comparative confidences are primitive.

in thinking we have an argument that each precise Bayesian is unreasonably proud. On the pluralist interpretation, the agent is undecided which of these unreasonable advisors she will follow. But such a state is itself unreasonable; she should have decided not to follow any of them, since they are all provably unreasonable!

A surprising fact about corporate bodies is that they can be immune to problems that beset each of their members. It would be illegitimate for any one parliamentarian to have law-making power; it is (or at least can be) legitimate for them all to have such power. Indeed, it would be unreasonable for any of them to think that they individually should have law-making powers; that would be unreasonably proud. But it is not unreasonable for them to collectively think that they should collectively have law-making powers. If they are a well-constituted parliament, this is a perfectly reasonable thought. Similarly here, the agent, the corporate body, could avoid being unreasonably proud even though each of the representors is over-confident in its own powers.

Now going back to success and modesty, it seems to me that the first definition of success is appropriate on the pluralist interpretation of the imprecise framework, and the second is appropriate on the corporate interpretation. The first interpretation says that the agent succeeds iff every member succeeds. And the second says that the agent succeeds iff the body of functions, collectively, succeed. Since I'm defending the use of the imprecise framework on the corporate interpretation, it is the second definition of success that is appropriate, and that's what I will use here.

This understanding isn't without costs. Bradley (2014) argues, in effect, that the best responses to dilation-based arguments against imprecise probabilities (as in White (2010)), are only available on the pluralist interpretation. I'm not going to try to solve those problems here, but I will note that the interpretative choice I'm making generates some extra philosophical work elsewhere. Against that, the corporate interpretation has some benefits. It lets us agree with Peter Walley (1991) that there are rational agents who are represented by sets of merely finitely additive probability functions, though no merely finitely additive probability function on its own could represent a rational agent. So the issues between the two interpretations are extensive. For now, I'll simply note that I'm interested in defending the imprecise Bayesian from Belot's argument on the corporate interpretation. And with that I'll return to translating Belot's puzzle into the imprecise framework, with the second, corporate-friendly, interpretation of success on board.

There are also two natural ways to generalise Belot's notion of open-mindedness to the imprecise case. We could require that the agent satisfies either the first or second of these conditions.

1.  $\forall \mathbf{x}_k \exists \mathbf{x}_i \supset \mathbf{x}_k, \mathbf{x}_j \supset \mathbf{x}_k: \neg(\Pr(p \mid \mathbf{X}_i = \mathbf{x}_i) \geq 1/2) \wedge \neg(\Pr(p \mid \mathbf{X}_j = \mathbf{x}_j) < 1/2)$
2.  $\forall \mathbf{x}_k \exists \mathbf{x}_i \supset \mathbf{x}_k, \mathbf{x}_j \supset \mathbf{x}_k: \Pr(p \mid \mathbf{X}_i = \mathbf{x}_i) < 1/2 \wedge \Pr(p \mid \mathbf{X}_j = \mathbf{x}_j) \geq 1/2$

The second is just the same symbols as in Belot's, and it is what I'll end up arguing is the right constraint to put on the imprecise Bayesian agent. And it is a considerably more demanding constraint than the first. But the first is perhaps the more natural understanding of *open-mindedness*. It says that no matter what the initial evidence is, the agent is not guaranteed to settle her credence in  $p$  on one side of  $\frac{1}{2}$ . That's a way of being open-minded.

But if the agent satisfies that constraint, she may be open-minded, but she won't necessarily be responsive to the evidence. Here's how I'm using the terms 'open-minded' and 'evidence-responsive'. In both clauses, the quantification is intended to be over a salient class of propositions. (The relevant class in the application we're most interested in is just  $\{X \text{ is periodic}, *X \text{ is not periodic}\}$ .) And I'll say an agent is 'confident' in a proposition iff her credence in it is above  $\frac{1}{2}$ .

**Open-Minded** Any time an agent is confident in a proposition, there is some evidence she could get that would make her lose confidence in it.

**Evidence-Responsive** For any proposition, there is some evidence the agent could get that would make her confident in it.

Once we allow imprecise credences, these two notions can come apart. Consider the agent we described above, whose representor consists of equal mixtures of the fair-coin measure and regular periodic functions. They are open-minded; they can always lose confidence that  $X$  is periodic or not. But they aren't evidence-responsive; no matter what the evidence, their credence that  $X$  is periodic will never rise above  $\frac{1}{2}$ . In fact, their credence that  $X$  is periodic will never rise above any positive number.

That suggests open-mindedness is too weak a constraint. If the evidence the agent gets is a string of several hundred 0s, she shouldn't just lose any initial confidence in  $\neg p$ , she should become confident in  $p$ . And arguably (though I could imagine a dissent here), if the initial sequence is a seemingly random sequence, the credence in  $p$  should drop well below  $\frac{1}{2}$ . (The imagined dissent here is from someone who thinks that the noisier the data, the more imprecise credences should get. That's an interesting view, but perhaps orthogonal to the issues we're debating here.)

And when we look back at Belot's motivations for open-mindedness, we see that they are really motivations for being evidence-responsive. One of the distinctive (and I would say problematic) features of precise Bayesianism is that it doesn't really have a good way of representing a state of indecisiveness or open-mindedness. In the terms we've been using here, there's no difference for the precise Bayesian between being evidence responsive and open minded. The imprecise Bayesian can distinguish these. And in Belot's puzzle, we should require that the imprecise Bayesian agent is evidence responsive. So we should impose the second, stronger, condition.

The final condition to discuss is modesty. There are three natural candidates here. We could merely require that the agent's prior probability that  $\mathbf{x}$  is in her success set is not equal to 1. Or we could require that it be less than 1. Or, even more strongly, we could require that it be less than some number that is less than 1. If her credence that  $\mathbf{x}$  is in her success set is imprecise over some interval  $[k, 1]$ , she satisfies the first condition, but not the second or third. If it is imprecise over some interval  $(k, 1)$ , or  $[k, 1)$ , she satisfies the first and second conditions, but not the third. In the interests of setting the imprecise Bayesian the hardest possible challenge, though, let's say that modesty requires the third criteria. Her ex ante credence in success should not just be less than 1, it should be less than some number less than 1.

The aim of the next section is to describe a representor that satisfies open-mindedness and modesty with respect to the question of whether the sequence is periodic. The representor will not represent a state that it is rational for a person to be in; we'll come back in the last section to the significance of this. My aim is just to show that for the imprecise Bayesian, unlike the precise Bayesian, open-mindedness and modesty are compatible. And the proof of this will be constructive; I'll build a representor that is, while flawed in some other ways, open-minded and modest.

### 3 Meeting the Challenge, Imprecisely

Recall that  $\text{Pr}_0$  is the *fair-coin measure*, according to which, if  $\mathbf{y}$  is any  $k$  length sequence of 0s and 1s we have  $\text{Pr}_0(\mathbf{X}_k = \mathbf{y}) = 2^{-k}$ .

Say a finite sequence  $\mathbf{y}_k$  of length  $k$  is **repeating** iff for some  $n > 1$ ,  $\mathbf{y}_k$  consists of  $n$  repetitions of a sequence of length  $k/n$ . For any non-repeating sequence  $\mathbf{y}_k$  (of length  $k$ ) let  $s_{\mathbf{y}_k}$  be the sequence consisting of  $\mathbf{y}_k$  repeated infinitely often. Let  $\text{Pr}_1$  be the function such that,

$$\text{Pr}_1(X = s_{\mathbf{y}_k}) = \frac{1}{2^{2k} - 1}$$

Intuitively, we can think of  $\text{Pr}_1$  as follows. Consider a measure over representations of periodic sequences. Any periodic sequence can be represented just as a finite sequence, plus the instruction *repeat infinitely often*, so this is really just a measure over finite sequences. One natural such measure assigns measure  $2^{-2k}$  to each sequence of length  $k$ . Of course, several of these representations will be representations of the same sequence. For instance,  $\langle 0, 1 \rangle$ ,  $\langle 0, 1, 0, 1 \rangle$  and  $\langle 0, 1, 0, 1, 0, 1 \rangle$  repeated infinitely produce the same sequence. Now the probability of a sequence, according to  $\text{Pr}_1$  is just the measure, so defined, of the class of representations of that measure. (It's a little easier to confirm that the measures sum to 1 than that the probabilities do, which is why I've

included this little explanation.)

Now define  $\text{Pr}_2$  as the equal weight mixture of  $\text{Pr}_0$  and  $\text{Pr}_1$ , i.e.,  $\text{Pr}_2(q) = (\text{Pr}_0(q) + \text{Pr}_1(q))/2$ . Since  $\text{Pr}_0(p) = 0$ , and  $\text{Pr}_1(p) = 1$ ,  $\text{Pr}_2(p) = 1/2$ . There will be several facts about  $\text{Pr}_2$  that are useful to have in place for future reference. (Recall I'm using  $\mathbf{X}$  as a random variable for the sequence the agent will see,  $\mathbf{x}$  as a rigid designator of that sequence,  $\mathbf{y}$  and  $\mathbf{z}$  are variables for arbitrary sequences, and the  $k$  subscript to restrict sequences to length  $k$ .) The first of these was proven as Lemma 1.

**Lemma 1.**  $\text{Pr}_2(p \mid \mathbf{X}_k = \mathbf{y}_k) > 1/2$  iff  $\text{Pr}_1(\mathbf{X}_k = \mathbf{y}_k) > \text{Pr}_0(\mathbf{X}_k = \mathbf{y}_k)$ .

Define a new predicate  $n$  of finite sequences  $\mathbf{y}_k$ , to hold just in case  $\mathbf{y}_k$  could be the initial segment of an infinite sequence of period at most  $k/2$ . So  $\mathbf{y}_k$  must consist of some sequence repeated twice, and anything else in  $\mathbf{y}_k$  must be consistent with that sequence repeating again (and if necessary again, and again, ...). Then we get,

**Lemma 2.** For  $k \geq 2$ ,  $\text{Pr}_2(p \mid \mathbf{X}_{2k} = \mathbf{y}_{2k}) > 1/2$  iff  $N\mathbf{y}_{2k}$ .

*Proof.* By Lemma 1, this reduces to the question of the relationship  $\text{Pr}_1(\mathbf{X}_{2k} = \mathbf{y}_{2k}) > \text{Pr}_0(\mathbf{X}_{2k} = \mathbf{y}_{2k})$ . Moreover, we know that  $\text{Pr}_0(\mathbf{X}_{2k} = \mathbf{y}_{2k}) = 2^{-2k}$ . So the question is whether  $\text{Pr}_1(\mathbf{X}_{2k} = \mathbf{y}_{2k}) > 2^{-2k}$ .

If  $N\mathbf{y}_{2k}$ , then it is consistent with  $\mathbf{X}_{2k} = \mathbf{y}_{2k}$  that  $\mathbf{x}$  is a particular periodic sequence with period at most  $k$ . Since the probability, according to  $\text{Pr}_1$  of any such sequence is greater than  $2^{-2k}$ , the right-to-left direction follows.

If  $\neg N\mathbf{y}_{2k}$ , then the possibilities that get positive probability according to  $\text{Pr}_1$  are at most among the following:  $\mathbf{X}$  consists of the first  $k + 1$  digits of  $\mathbf{y}_{2k}$  repeated endlessly;  $\mathbf{X}$  consists of the first  $k + 2$  digits of  $\mathbf{y}_{2k}$  repeated endlessly; ...;  $\mathbf{x}$  consists of the first  $2k$  digits of  $\mathbf{y}_{2k}$  repeated endlessly;  $\mathbf{X}$  is one of the two sequences of period  $2k + 1$  starting with  $\mathbf{y}_{2k}$ , or one of the four sequences of period  $2k + 2$  starting with  $\mathbf{y}_{2k}$  or .... So we get the following, starting with the probabilities of each of the possibilities listed in the previous sentence,

$$\begin{aligned} \text{Pr}_1(\mathbf{X}_{2k} = \mathbf{y}_{2k}) &\leq \frac{1}{2^{2k+2} - 1} + \frac{1}{2^{2k+4} - 1} + \dots + \frac{1}{2^{4k} - 1} + \frac{2}{2^{4k+2} - 1} + \dots \\ &< \frac{1}{2^{2k+1}} + \frac{1}{2^{2k+3}} + \dots + \frac{1}{2^{4k-1}} + \frac{1}{2^{4k}} + \dots \\ &< \frac{1}{2^{2k}} \end{aligned}$$

And from that the left-to-right direction follows.  $\square$

**Lemma 3**  $\text{Pr}_2$  is open-minded.

*Proof.* Since any initial sequence  $\mathbf{y}_k$  that is not  $N$  can be easily extended into one that is  $N$  (by, e.g., repeating  $\mathbf{y}_k$ ), and one that is  $N$  can be extended into one that is not (by, e.g., having the repeating sequence stop at the very next step), this follows immediately from Lemma 2.  $\square$

Define  $f$  to be a function from sequences of length  $k \geq 2$  to sequences of length  $k + 1$  such that

$$f(y_k) = y_k + \begin{cases} \langle 0 \rangle & \text{if } Ny_k \leftrightarrow \text{Pr}_1(x_{k+1} = 0 | X_k = y_k) \leq \frac{1}{2} \\ \langle 1 \rangle & \text{otherwise} \end{cases}$$

In the normal way, define  $f^n(\mathbf{y}_k)$  to be the result of applying  $f$   $n$  times to  $\mathbf{y}_k$ . And define  $f^\infty(\mathbf{y}_k)$  to be the infinite sequence we get by doing this infinitely often.

Intuitively, the way  $f$  works is that if  $\mathbf{y}_k$  is already somewhat sequential, then we include the less likely digit, and if it isn't, then we include the more likely digit. (With ties resolved in favour of including 0 rather than 1.) If we define  $p(\mathbf{y}_k)$  to be the smallest  $n$  such that  $\mathbf{y}_k$  could be the initial segment of a periodic sequence of length  $n$ , then we'll get that  $p(f(\mathbf{y}_k)) > p(\mathbf{y}_k) \leftrightarrow Ny_k$  in all cases, except for the case where  $\text{Pr}_1(\mathbf{x}_k = 0 | \mathbf{X}_k = \mathbf{y}_k) = \frac{1}{2}$ . That is, if  $Ny_k$ , then extending  $\mathbf{y}_k$  in this way will wipe out the possibility of that smallest sequence being extended indefinitely, while if  $\neg Ny_k$ , then that possibility will still be on the table.

From this, it follows that  $f^\infty(\mathbf{y}_k)$  will flummox  $\text{Pr}_2$ , no matter which  $\mathbf{y}_k$  we start with.

We need one last classification of finite sequences, and then we are done. Say that  $Oy_k$  just in case some initial segment of  $\mathbf{y}_k$  of length  $r$  could be the initial segment of an infinite period sequence of period less than  $r/2$ . This contrasts with  $N$  in two ways. First, it requires a sequence that repeats twice, and then starts a third repetition. Second, it does not require that the sequence be 'live'; there might be subsequent parts of  $\mathbf{y}_k$  that are not compatible with the sequence repeating. So the sequence  $\langle 0, 0, 1, 0, 0, 1 \rangle$  satisfies  $N$  but not  $O$ , while the sequence  $\langle 0, 1, 0, 1, 0, 0 \rangle$  satisfies  $O$  but not  $N$ .

There are a countable infinity of finite sequences  $\mathbf{y}_k$  such that  $\neg Oy_k$ . Produce some ordering of them, then define  $\text{Pr}_i$ , for  $i \geq 3$ , to be the probability function such that  $\text{Pr}_i(\mathbf{X} = f^\infty(\mathbf{y}_k)) = 1$ , where  $\mathbf{y}_k$  is the  $i-2$ 'th sequence in this order.

Now, consider the set  $R$  of all probability functions of the form:

$$\text{Pr} = \sum_{i=2}^{\infty} a_i \text{Pr}_i$$

where each of the  $\text{Pr}_i$  are defined as above, each  $a_i$  is non-negative,  $a_2$  is  $\frac{1}{2}$ , and the sum of the  $a_i$  from 3 to  $\infty$  is also  $\frac{1}{2}$ . Intuitively, each function starts by halving the

probability  $\text{Pr}_2$  gives to each initial (or completed) sequence, and distributing the remaining probability over the countable infinity of flummoxing sequences of the form  $f^\infty(\mathbf{y}_k)$ , where  $\neg O\mathbf{y}_k$ .

I'll now prove that  $R$  is open minded.

**Lemma 4** If  $\neg O\mathbf{y}_k$ , then  $\neg Of(\mathbf{y}_k)$ .

*Proof.* Since  $\neg O\mathbf{y}_k$ , the only way that  $Of(\mathbf{y}_k)$  could be true is if  $k = 2r + 1$ , and  $f(\mathbf{y}_k)$  consists of some sequence of length  $r$  repeated twice, plus the first digit repeated a third time. But that means that  $N\mathbf{y}_k$ . And if that's the case, then the extra digit that is added by  $f(\mathbf{y}_k)$  will not be the necessary digit to repeat this sequence. So it is impossible that  $Of(\mathbf{y}_k)$ .  $\square$

**Lemma 5** If  $\neg O\mathbf{y}_k$ , then  $\neg Of^\infty(\mathbf{y}_k)$ .

*Proof.* This follows trivially from Lemma 4.  $\square$

**Theorem 6**  $R$  is open-minded.

*Proof.* Any initial sequence can be extended to a sequence satisfying  $O$ . For example, the initial sequence can be repeated in full twice. An immediate consequence of Lemma 5 is that for all  $i \geq 3$ ,  $O\mathbf{y}_k \rightarrow \text{Pr}_i(\mathbf{X}_k = \mathbf{y}_k) = 0$ . That means that if  $O\mathbf{y}_k$ , then for any  $\text{Pr} \in R$ ,  $\text{Pr}(p \mid \mathbf{X}_k = \mathbf{y}_k) = \text{Pr}_2(p \mid \mathbf{X}_k = \mathbf{y}_k)$ . And now the theorem is an immediate consequence of Lemma 3.  $\square$

Let  $F$  be the set of all sequences  $f^\infty(\mathbf{y}_k)$ , where  $\neg O\mathbf{y}_k$ .

**Lemma 7** If  $\mathbf{x} \in F$ , then  $R$  fails.

*Proof.* Assume  $\mathbf{x} \in F$ , so  $\mathbf{x}$  is not periodic. Then proving the lemma requires showing that for any  $i$ , there is a  $j \geq i$  such that, according to  $R$ , the probability of  $p$  given  $\mathbf{X}_j = \mathbf{x}_j$  is not less than  $1/2$ . And that requires showing that there is a  $\text{Pr} \in R$  such that  $\text{Pr}(p \mid \mathbf{X}_j = \mathbf{x}_j) \geq 1/2$ . This is easy to do. Consider any sequence  $\mathbf{y}_i$  of length  $i$  not identical to  $\mathbf{x}_i$  such that  $\neg O\mathbf{y}_i$ . Consider the probability function  $\text{Pr}_k \in R$  such that  $\text{Pr}_k(\mathbf{X} = f^\infty(\mathbf{y}_i)) = 1/2$ . Once we conditionalise on  $\mathbf{X}_i = \mathbf{x}_i$ , that function will behave just like  $\text{Pr}_2$ . And since  $\mathbf{X}$  flummoxes  $\text{Pr}_2$ , that means there is a  $\mathbf{x}_j$  such that  $\text{Pr}(p \mid \mathbf{X}_j = \mathbf{x}_j) > 1/2$ , and hence  $\text{Pr}(p \mid \mathbf{X}_j = \mathbf{x}_j) \geq 1/2$ .  $\square$

**Lemma 8** For each  $\text{Pr} \in R$ ,  $\text{Pr}(\mathbf{x} \in F) = 1/2$ .

*Proof.* It helps to think of each of the  $\text{Pr} \in R$  as mixtures of  $\text{Pr}_0$  and  $\text{Pr}_1$ , plus a mixture of the  $\text{Pr}_i$  for  $i \geq 3$ . Now  $\text{Pr}_0(\mathbf{x} \in F) = 0$ , since for any countable set,  $\text{Pr}_0$  says the probability that  $\mathbf{x}$  is in that set is 0. And  $\text{Pr}_1(\mathbf{x} \in F) = 0$ , since  $\text{Pr}_1$  says that the probability of  $\mathbf{x}$  being periodic is 1, and none of the members of  $F$  are periodic. But for each  $\text{Pr}_i$  for  $i \geq 3$ ,  $\text{Pr}_i(\mathbf{x} \in F) = 1$ . Indeed, for each such function, there is a particular sequence in  $F$  such that the probability that  $\mathbf{x}$  is that sequence is 1. So for each  $\text{Pr} \in R$ ,  $\text{Pr}(\mathbf{x} \in F) = \frac{1}{4} \times 0 + \frac{1}{4} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}$ .  $\square$

**Theorem 9** According to  $R$ , the probability of an agent whose representor is  $R$  failing is at least  $\frac{1}{2}$ .

*Proof.* Immediate from Lemma 7 and Lemma 8.  $\square$

So if an agent's credences are represented by a non-singleton set of probability functions, not a single probability function, it is possible for them to be open-minded and modest. On the other hand, if an agent is represented by a single probability function, as the precise Bayesian desires, then it is impossible to be open-minded and modest. Since being open-minded and modest is desirable, this is a reason to prefer the imprecise Bayesian picture.

## 4 Objections and Replies

I'm going to reply to three objections, but since my replies overlap, I'll group the objections together.

**Objection 1** The model here only gives you conditional modesty. Once the initial sequence is  $O$ , the representor becomes the singleton of an open-minded probability function, and Belot showed that to be immodest. Ideally, the agent would have a prior that is in some way resiliently modest, whereas this prior is fragilely modest.

**Objection 2** This representor is open-minded and modest towards one particular problem, namely whether  $\mathbf{X}$  is periodic. But Belot was interested in a wider range of problems, indeed in all problems of the form: does  $\mathbf{x}$  fall into some set that is measurable, dense, and has a dense complement. Ideally, we'd have a prior which is widely open-minded and modest, in the sense that it had an open-minded and modest attitude towards many problems. But this prior is narrowly modest, in the sense that it is open-minded and modest about only one problem.

**Objection 3** The representor described here is clearly not a representation of a credal state of anyone rational. Look what it does if the data is a 1 followed by thousands of 0s, or is the first few thousand digits of the binary expansion of  $\pi$ , or has a frequency of 0s of 0.2 over all large sub-intervals. No one could adopt this prior, so it doesn't show anything about the advantages of imprecise Bayesianism.

*Reply.* My responses are going to be (1) that we should want more resilient modesty, and though this is a hard technical challenge, it's possible to see a way forward on it, (2) that we should want somewhat wider open-minded modesty, though how much wider is a hard question, and (3) that the third objection should simply be rejected. Let's go through those in reverse order, since it's the response to the third that explains part of what I'm doing in response to the other two.

What we have in section three is a consistency proof. For the imprecise Bayesian, unlike the precise Bayesian, being open-minded is consistent with being modest. That's good, since it shows that we can't rule out a rational response to problems like Belot's. It's obviously true that the prior in question isn't rational, but that's not needed for a consistency proof.

Moreover, we don't just have a consistency proof, we have a constructive consistency proof - the prior is described in detail. It's just not going to be possible to do a constructive proof that open-mindedness, modesty and full rationality are consistent. And that's because to do that would essentially be to solve all of the problems of epistemology ever. Demonstrating a fully rational prior, even for the range of questions Belot considers, is too much to ask.

If there's a reasonable looking argument that imprecise Bayesians are unlikely to be able to satisfy some set of plausible constraints, then the defender of imprecise Bayesianism is, I think, obliged to show how those constraints can be satisfied. But to ask for a demonstration of how all reasonable constraints can be satisfied at once, in the absence of a decent argument that they cannot be, would clearly be asking too much.

So I don't care that the prior I described is irrational; it serves its purpose in proving consistency. Now what would be nice is to show that some slightly stronger constraints can be simultaneously satisfied. But we have to be sure that those constraints are in fact reasonable constraints. Here's one constraint that I think isn't reasonable: be open-minded towards any proposition of the form  $\{\mathbf{X} \in S\}$ , where  $S$  is a dense set of sequences. Let  $S$ , for example, be the set consisting of all sequences of the form  $\mathbf{y}_k + \mathbf{z}$ , where  $\mathbf{y}_k$  ranges over all finite sequences, and  $\mathbf{z}$  is a particular arbitrary sequence that lacks finite definition in our current language. That set is dense, and indeed measurable. But there's no evidence that could make it reasonable to take  $\mathbf{X} \in S$  to be probable. So a prior that wasn't open-minded towards  $\mathbf{X} \in S$  could still be perfectly reasonable.

That said, the prior I demonstrated is closed-minded towards several propositions that should be taken seriously. It will never have positive credence that  $\mathbf{X}$  is eventually periodic without being periodic, or that  $\mathbf{X}$  is generated by a chance process that gives each data point chance  $c \neq \frac{1}{2}$  of being 0. It would be good to have a prior whose open-minded modesty was wider. But before we do that technical work, I think there's a need to figure out which propositions we should be open-minded about.

I am more worried by the fragility of the modesty of this prior. There's a reasonable sense in which the prior is open-minded only in virtue of the fact that it has parts which are immodest. At any point where the agent has credence above  $\frac{1}{2}$  that  $p$ , she has credence 1 that she will succeed.

We could try to complicate the prior a bit more to avoid that. Here's a sketch of how it could go, with application to one particular initial sequence of data. Consider what happens to  $R$  if the initial input is  $\langle 0, 1, 0, 0, 1, 0, 0, 1 \rangle$ , hereafter  $\mathbf{y}$ . According to  $\text{Pr}_0$ , that initial sequence has probability  $1/256$ . According to  $\text{Pr}_1$ , it has probability  $1/63 + 1/4095 + 1/65535 \approx 1/62$ . So given that initial sequence,  $\text{Pr}_2$  says the probability of  $p$  is about 0.8. And since the sequence is  $O$ , it could be the start of the the sequence  $\langle 0, 1, 0 \rangle$  repeated indefinitely, its probability according to  $\text{Pr}_i$  is 0, for  $i \geq 3$ . Now consider the set of all probability functions of the form  $a\text{Pr}_R + b\text{Pr}_{New}$ , where  $a + b = 1$ ,  $b \in (0, 1/256)$ ,  $\text{Pr}_R \in R$  and  $\text{Pr}_{New}$  is the function which gives probability 1 to  $\mathbf{X}$  being  $O(\mathbf{y})$ . That prior is open-minded, and even after conditionalising on  $\mathbf{y}$  satisfies the intermediate of the three modesty conditions described on page - the probability of failure is less than one, though it isn't less than some number less than one. And this trick could be generalised to satisfy more modesty conditions, and even (though it would take some time to prove this) be unconditionally modest.

But I'm not going to go through those steps here. That's mostly because I think we already have shown enough to show that imprecise Bayesianism has an advantage over precise Bayesianism. The imprecise Bayesian can, and the precise Bayesian can't, have an open-minded modest attitude. It would be good to press home that advantage and show that there are other things the imprecise Bayesian can do that the precise Bayesian can't do, such as having a widely open-minded and resiliently modest prior. But even before such a demonstration takes place, the advantage has been established.

## References

- Belot, Gordon. 2013. "Bayesian Orgulity." *Philosophy of Science* 80 (4): 483–503. doi: 10.1086/673249.
- Bradley, Seamus. 2014. "Imprecise Probabilities." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2014. <http://plato.stanford.edu/archives/win2014/entries/imprecise-probabilities/>; Meta-

physics Research Lab, Stanford University.

Briggs, Ray. 2009. "Distorted Reflection." *Philosophical Review* 118 (1): 59–85. doi: 10.1215/00318108-2008-029.

Elga, Adam. 2010. "How to Disagree about How to Disagree." In *Disagreement*, edited by Ted Warfield and Richard Feldman, 175–87. Oxford: Oxford University Press.

Joyce, James M. 2010. "A Defence of Imprecise Credences in Inference and Decision Making." *Philosophical Perspectives* 24 (1): 281–323. doi: 10.1111/j.1520-8583.2010.00194.x.

Lasonen-Aarnio, Maria. 2015. "New Rational Reflection and Internalism about Rationality." *Oxford Studies in Epistemology* 5: 145–71. doi: 10.1093/acprof:oso/9780198722762.003

Lewis, David. 1971. "Immodest Inductive Methods." *Philosophy of Science* 38 (1): 54–63. doi: 10.1086/288339.

———. 1993. "Many, but Almost One." In *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*, edited by Keith Campbell, John Bacon, and Lloyd Reinhardt, 23–38. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511625343.010. Reprinted in his *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 1999, 164–182. References to reprint.

Pryor, James. 2000. "The Sceptic and the Dogmatist." *Noûs* 34 (4): 517–49. doi: 10.1111/0029-4624.00277.

Schoenfield, Miriam. 2012. "Chilling Out on Epistemic Rationality: A Defense of Imprecise Credences (and Other Imprecise Doxastic Attitudes)." *Philosophical Studies* 158 (2): 197–219. doi: 10.1007/s11098-012-9886-7.

———. 2015. "A Dilemma for Calibrationism." *Philosophy and Phenomenological Research* 91 (2): 425–55. doi: 10.1111/phpr.12125.

Walley, Peter. 1991. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman & Hall.

White, Roger. 2006. "Problems for Dogmatism." *Philosophical Studies* 131 (3): 525–57. doi: 10.1007/s11098-004-7487-9.

———. 2010. "Evidential Symmetry and Mushy Credence." *Oxford Studies in Epistemology* 3: 161–89.

Published in *Ergo*, 2016, pp. 529–545.