Self-Locating Belief and Updating on Learning

DARREN BRADLEY

University of Leeds

d.j.bradley@leeds.ac.uk

1. Introduction

Beliefs that locate you in space or time are *self-locating beliefs*. These cause a problem for Bayesian models of belief. Miriam Schoenfield (2016) offers a solution – that on learning x, agents should update on the fact that they learned x. I will argue that Schoenfield's suggestion does not solve the problem.

2. Background: Conditionalization and Self-Locating Belief

Let's start with the Bayesian picture of belief update. Imagine all the possible worlds spread out across logical space. Each has some initial probability. When you learn E, all the not E possibilities are eliminated, and their probabilities are distributed across the remaining E possibilities. This process of eliminating false possibilities and zooming in on the one true world continues until, at the limit of enquiry, when omniscience is reached, there is only one possibility remaining – the one true world has all the credence and there is no more learning to be done.

But now consider a self-locating belief such as *today is Sunday*. Suppose you acquire omniscience on a Sunday, and so believe that it is Sunday. The problem is that if you stick with this belief for 24 hours you belief falsely – it's now Monday but you falsely believe it's Sunday. If you want to stay omniscient you need to give up the belief that it's Sunday and acquire the belief that it's Monday. But traditional Bayesianism simply doesn't have a mechanism for this kind of belief update. The problem is that self-locating facts are a moving target. There is no one true world to zoom in on. Instead, the true world is a Sunday world, then a Monday world, then a Tuesday world…

To put the point formally, the standard model of belief update for Bayesians is conditionalization:

Conditionalization: If an agent learns E and nothing else between t0 and t1 then

P1(.) = P0(.|E)

Now consider a self-locating belief such as *today is Sunday.* You might be certain that it is Sunday, and hear the clock has struck midnight, meaning it's now Monday. Conditionalization seems to imply that you should believe that it is Monday, and *maintain* your belief that it is Sunday i.e. believe it is Sunday and Monday. But clearly this is not what you should believe.

This suggests that the problem is due to certainty, and that we might solve it by adopting a model in which agents are never certain. But certainty is not essential to the problem.[1] The core of the problem is that self-locating beliefs can be true at one time and false at a later time (and vice versa); yet the Bayesian model is that of agents eliminating possibilities and zooming in on the truth.

Miriam Schoenfield (2016) defends an alternative to conditionalization and argues that it solves the problem of self-locating belief. I will argue that it cannot solve the problem.

3. Schoenfield's Theory

Here is Schoenfield's suggestion (we grant with Schoenfield that one should maximize expected accuracy):

> I will now argue that conditionalizing on self-locating evidence doesn't in general maximize expected accuracy in cases of belief discovery[2]…I will show that the update procedure that does, in general, maximize expected accuracy in cases of

[1] See Hedden (2015 p.37-40)

[2] Schoenfield holds that the reason they shouldn't conditionalize on self-locating belief is that 'Factivity' doesn't hold i.e. the agent shouldn't conditionalize because the agent is not certain at t0 that if she learns E upon undergoing her future learning experience [at t1], then E is true at t0. But this means it's not a case of Discovery (see Bradley 2011a p.395); Discovery says that the agent *is* certain that the truth-value of E does not change over the period of interest. I maintain that in cases of belief Discovery, conditionalizing on self-locating evidence does maximize expected accuracy; if Factivity fails then it is not a case of Discovery.

> self-locating evidence is not conditionalization, but what I will call conditionalization* [the update procedure that has us adopt p(.|L(Xi)), upon learning Xi]. 2016 p.702

> Generalized CondMax: Suppose that you are certain that you are going to learn exactly one proposition from a set of propositions, X, at time t. Let L(Xi) be the proposition that Xi is learned upon undergoing the learning experience at t. The update-procedure that maximizes expected accuracy in response to X, relative to probability function p, is the update-procedure that assigns, to each Xi, P(.|L(Xi)). 2016 p.703[3]

Schoenfield isn't explicit about whether L(Xi) is intended to be a self-locating proposition or a non-self-locating proposition. Is it the self-locating proposition that Xi is learnt *now,* or is it the non-self-locating proposition that Xi is learnt at some time? Either way, it seems to run into a problem, so my challenge takes the form of a dilemma.

4. First horn: L(Xi) is self-locating

On the first horn, suppose L(Xi) is self-locating. This horn fails to avoid the original problems caused by self-locating beliefs. Consider an agent who is certain that it is Sunday and then acquires the belief that it is Monday as time passes. Then Xi = Today is Monday. The original problem was that conditionalizing on 'Today is Monday' results in the agent absurdly believing 'Today is Sunday and today is Monday'; the earlier belief that today is Sunday has not gone away. On Schoenfield's account the agent should not conditionalize on 'Today is Monday' – they should conditionalize on 'That today

[3] This is developed in Schoenfield 2017. I am broadly in agreement with Schoenfield, but I think that in trying to solve the problem of self-locating beliefs, she overreaches.

is Monday is learned'. And to make explicit that this is self-locating, we can put it as 'That today is Monday is learned today'.

But this doesn't address the problem. The agent can trivially infer from 'That today is Monday is learned today' that today is Monday, and is once again left with the absurd belief that it is Sunday and Monday. The agent's earlier belief that today is Sunday has not gone away. The absurd result that they believe 'today is Sunday and today is Monday' remains.

5. Second horn: L(Xi) is non-self-locating

On the second horn, suppose L(Xi) is non-self-locating. This means the procedure Schoenfield recommends is to conditionalize on the non-self-locating proposition that Xi is learnt at some time. But there are cases where conditionalizing on only non-self-locating propositions gives the wrong answer. Indeed, Schoenfield herself explains the point with the following case:

> FAIRIES AND DEMONS: Sleeping Beauty is going to be awoken and put back to sleep ten times, beginning on Monday morning. After each awakening, she will encounter either a friendly fairy or an evil demon and then her memory of the awakening will be erased before she is put back to sleep. How many fairies or demons she will see will be determined by whether she was blessed or cursed at the time of her birth. If she was blessed, she'll see a friendly fairy on nine days and an evil demon on one day, but if she was cursed she'll see an evil demon on nine days and a friendly fairy on one day. Beauty knows all of this before going to sleep and her initial credence that she was blessed at the time of her birth is 0.5. (2016 p.11)

Schoenfield agrees that seeing a fairy confirms that Beauty is blessed (and seeing a demon confirms that Beauty is cursed). Obtaining such confirmation is a desired result of any analysis. And Schoenfield concedes that:

> [conditionalizing on non-self-locating] propositions like *Beauty sees a fairy on one of the awakenings…*won't yield the desired result. Since Beauty is certain that…she will see a fairy on one of the awakenings, that she sees a friendly fairy on one of the awakenings provides no evidence either for being blessed or for being cursed. p.11

And Schoenfield cannot deliver the desired result. This horn allows Beauty to conditionalize only on something like the non-self-locating 'Beauty learns that she sees a fairy on one of the awakenings', which fails to confirm that she is blessed. To get the desired result, Beauty needs to conditionalize on a *self-locating* piece of evidence such as 'I see a fairy *today*'. But self-locating evidence takes us back to the first horn of the dilemma. So Generalized CondMax does not avoid the original problem with self-locating evidence.

A referee suggests a response:

> Schoenfield may have a response along the following lines: The view is that L(X) is a non-self-locating proposition of the form "I learn X at time t" (or "X is learned at time t"). t can refer to a time demonstratively. In the Fairies and Demons case, even if I don't know what day it is when I wake up, before discovering whether there's a fairy or demon, I can name the time at which I will learn this information - say I name it "t". It's true, then, that prior to seeing the fairy on that particular day, Pr(I was blessed|I see a fairy at time t) = 0.9.[4]

[4] This is related to Titelbaum's (2008) theory. For elaboration on my response see Bradley (2011b p.333).

This is a tempting thought but I don't think it can work. Conditionalization requires that the same proposition which was uncertain at the earlier time is learnt at the later time. And there doesn't seem to be any reading of 't' which allows this.

Suppose 't' refers demonstratively, as the referee suggests, so the proposition is 'I see a fairy on this day'. The problem is that when uttered on earlier days, 'this day' refers to a different day. On Sunday, Pr(I was blessed|I see a fairy on this day) = 0.5, as 't' refers to Sunday, and seeing a fairy on Sunday is not relevant to being blessed or cursed. So applying conditionalization to 'I see a fairy on this day' on Monday, we get $Pr_{\text{I see a fairy on this day}}$(I was blessed) = 0.5. This is the wrong answer.[5]

Suppose 't' means 'whatever day, if any, that a fairy is seen'. But then the proposition learnt tells her nothing new i.e. Pr(I see a fairy whatever day, if any, that a fairy is seen) = 1, so Pr(I was blessed|I see a fairy whatever day, if any, that a fairy is seen) = 0.5. So I don't think the referee's response can work.

Stepping back, where does Schoenfield's argument for Generalized CondMax go wrong? Her argument (in her 2016 appendix 2) is based on Greaves and Wallace's (2006) argument for conditionalization, and Greaves and Wallace do not consider self-locating evidence. So self-locating evidence seems to introduce complexities that are not taken into account in Schoenfield's discussion.

6. Conclusion

Schoenfield argues that updating on 'I learn E' solves the problem of self-locating evidence. I argued that 'I learn E' must be either self-locating or not, and in each case fails to solve the problem.[6]

References

[5] We do get the right answer regarding the shift from being woken on one of the days to seeing a fairy on that same day, but this is because no self-locating proposition has changed truth-value over that period. We still need a theory for how beliefs should change between one day and the next.

[6] I am grateful to Miriam Schoenfield for helpful discussion and comments on an earlier draft.

Bradley (2011a) “Self-Location Is No Problem for Conditionalization.” Synthese 182: 393–411.

Bradley (2011b) ‘Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty’ *The British Journal for the Philosophy of Science*, Vol. 62, No. 2, pp.323-342

Greaves, H. and Wallace, D. (2006). ‘Justifying conditionalisation: conditionalisation maximizes expected epistemic utility’, *Mind* 115(459): 607-32

Hedden, Brian (2015). *Reasons Without Persons: Rationality, Identity, and Time*. Oxford University Press UK

Schoenfield, Miriam (2016). An Accuracy Based Approach to Higher Order Evidence. 96(3): 690-715 *Philosophy and Phenomenological Research.*

Schoenfield, Miriam (2017). Conditionalization Does Not (in general) Maximize Expected Accuracy. *Mind* 126 (504):1155-1187.

Titelbaum, M. 2008. “The Relevance of Self-Locating Beliefs.” *Philosophical Review* 117: 555–606.