Synthese
DOI 10.1007/s11229-010-9748-9

Self-location is no problem for conditionalization

D. J. Bradley

Received: 3 January 2010 / Accepted: 5 May 2010

Abstract How do temporal and eternal beliefs interact? I argue that acquiring a temporal belief should have no effect on eternal beliefs for an important range of cases. Thus, I oppose the popular view that new norms of belief change must be introduced for cases where the only change is the passing of time. I defend this position from the purported counter-examples of the Prisoner and Sleeping Beauty. I distinguish two importantly different ways in which temporal beliefs can be acquired and draw some general conclusions about their impact on eternal beliefs.

Keywords Sleeping Beauty · Self-location · Conditionalization · The Prisoner

1 Introduction

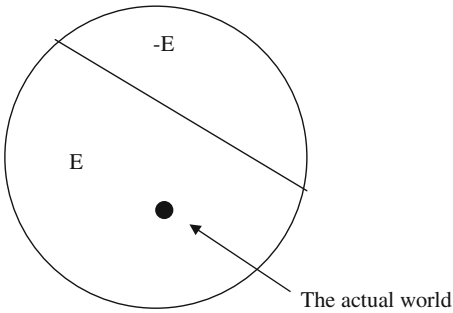
Sometimes we learn what the world is like. Other times, we learn about where we are in the world. This paper is about the way these two kinds of belief interact. I will argue that self-locating beliefs have less impact on non-self-locating beliefs than has been suggested in the recent literature. Theories which say they do have an impact are motivated by particular thought experiments, notably the Prisoner (Arntzenius 2003) and Sleeping Beauty (Elga 2000). I will argue that these cases have been mishandled and do not show any puzzling connections between self-locating and non-self-locating belief.

D. J. Bradley (✉)
The City College of New York, 160 Convent Avenue, New York, NY 10031, USA
e-mail: bradleydarren@gmail.com

Published online: 30 May 2010

2 Conditionalization and two types of belief change

Our most successful theory of confirmation—Bayesian confirmation theory—admits one basic rule of belief change: conditionalization. This says that an agent's degree of certainty, or credence, in a belief after learning a piece of evidence should equal their earlier degree of certainty in the belief, conditional on the evidence. Formally, if an agent has prior probabilities $P(H_i)$ at t0, and learns E and nothing else between t0 and t1, then her t1 probabilities should be $P(H_i|E)$, where $P(E) > 0$. More succinctly, $P_E(H_i) = P(H_i|E)$. Here is a useful way to think about conditionalization: A range of worlds are initially (epistemically) possible for the agent. One of these worlds is the actual world, the others are not. E is true in some worlds but not others, so the agent is uncertain about E. When the agent learns E, he eliminates all the not E worlds and increases his probability in the E worlds. Thus conditionalization can be pictured as the agent eliminating false possibilities and zooming in on the truth. Note that learning something that was previously uncertain is an essential part of conditionalization.



Traditional confirmation theory deals only with belief contents that do not change in truth-value over time. Call these eternal beliefs.[1] But many of our belief contents do change in truth-value over time.[2] Call these temporal beliefs.[3] Temporal beliefs are beliefs that locate the agent in time, such as 'it is 12:00'.[4] Such beliefs can be learnt in the way just described—by eliminating possibilities. For example, when someone is uncertain what time it is and looks at her watch, she acquires a new temporal belief by

[1] Beliefs that must be relativized to an agent will count as eternal for us e.g. I am DJB.

[2] I follow Kaplan (1989) in allowing that content can be temporal. But my arguments do not depend on this. In fact I am inclined to hold that contents are eternal (Schaffer ms) If you dislike temporal contents, replace my talk of contents with characters, roles or sentences. The sentence 'it is Monday' definitely changes in truth value.

[3] For ease of exposition, I will assume these categories, temporal and eternal, are mutually exclusive and exhaustive. We may learn both in virtue of the same experience. All I need is that any beliefs acquired can be divided into an eternal component and a temporal component.

[4] See Perry (1979), Lewis (1979). Temporal beliefs need not be explicitly about the time. For example the belief 'The sun is shining' will be temporal on our definition. It locates the agent at a time when the sun is shining.

eliminating false possibilities. (They are false relative to her current temporal position, not absolutely.)

In such cases, the temporal belief discovered does not change in truth-value over the period in which it changes from being disbelieved to being believed. Instead, the agent learns something of which they were initially uncertain. False possibilities have been eliminated, and the agent has 'zoomed in' on the truth, exactly as needed for conditionalization. Call this type of belief change Discovery.

> *Discovery*: Belief change in virtue of the discovery of the truth of the content of the belief, where the truth-value did not change over the period of interest.

But this story cannot always be applied to temporal beliefs. The reason is that temporal contents change in truth value. The content of 'it is the 21st century' has changed in truth value; it used to be false, but is now true. We need to change our temporal beliefs in accordance with the changing facts. Thus, we acquire temporal beliefs in virtue of the content of the belief changing from false to true.[5] (We also lose them when they change from true to false.) Call this type of belief change *Belief Mutation*[6].

> *Belief Mutation*: Belief change in virtue of a change in the truth-value of the content of the belief[7].

Belief Mutation applies only to cases where the agent learns only temporal information, such as staring at the hands of a clock in the knowledge that nothing unexpected is about to happen. By contrast, staring at the hands of a cuckoo clock at it strikes the hour may result in various eternal beliefs being learnt, for example that the cuckoo is red. Titelbaum (2008) gives the example of someone watching a film that they have seen so many times that they know every frame by heart (p. 556).

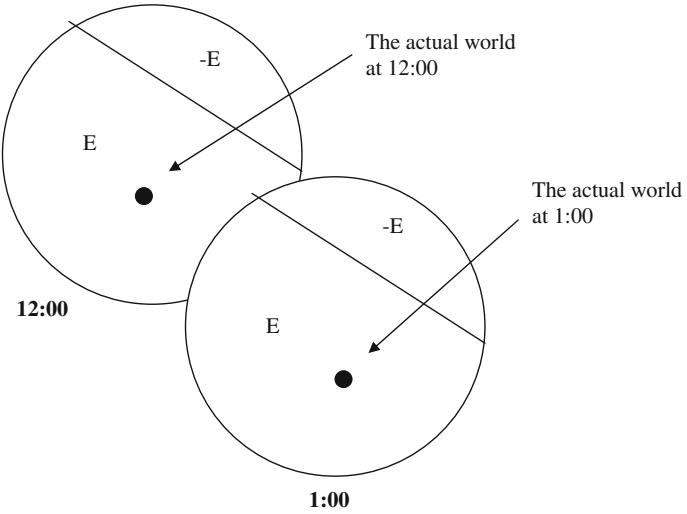
Belief Mutation has been overlooked until relatively recently, but once noted it is clear that it creates a problem for conditionalization being the *only* basic rule of belief change. Conditionalization says that beliefs should change only when something that was previously uncertain has been discovered. But Mutation allows belief change even when nothing that was uncertain is discovered. As I watch the hands of a clock move I may be uncertain about nothing, but my beliefs about the time should still change. And the reason they should change is that the truth value of the beliefs change. It used to be, say, 12:00 but now it is 1:00.[8]

[5] One complication I want to bracket is that we might be mistaken, so what really matters is whether we *think* the belief has changed in truth-value. I will assume this possibility of error won't affect my arguments. Note the same complication applies to Discovery and is generally ignored or assumed away.

[6] The change can usefully be described as mutation due to features it shares with biological mutation. It happens naturally over time, for example, and no involvement from other beliefs is necessary, just as no interference from other organisms is needed in biology.

[7] I will sometimes call this just 'Mutation'.

[8] I first saw diagrams similar to those that follow in a talk by Andy Egan in 2006.



Conditionalization is clearly the wrong model for this type of belief change, and quickly leads to absurdity if we try to apply it[9]. So when temporal beliefs are taken into account, conditionalization is not the only basic rule of belief update; we need new rules governing Mutation. To briefly give a sense of these rules, some will be simple such as 'If 24 hours pass, the belief [it is Monday today] should be replaced by [it was Monday yesterday]'. Similarly, 'If it stops raining, the belief [it is raining] should be replaced by [it was raining earlier]'. If the agent loses track of time they will be more complex, perhaps resulting in a weighted sum: 'if there is a 50% chance 24 hours have passed and a 50% chance 48 hours have passed, then replace the belief [it is Monday] by assigning a 50% chance to [it was Monday yesterday] and a 50% chance to [it was Monday two days ago]'. Mutation and Discovery may occur together; I will assume that all belief change can be decomposed into a Mutation component and a Discovery component. For example, if I wake up and see it is 7:52, then there is Mutation (I give up my last conscious belief that it was midnight) and also Discovery (I have discovered it is 7:52 rather than 7:51).[10]

But what happens to confirmation theory when temporal beliefs are admitted? That is, what is the relation between evidence and hypothesis when evidence and hypothesis can be temporal? Six possibilities need to be distinguished. Let's say (following Titelbaum) that if the evidence can shift one's degree of belief in the hypothesis,[11] the evidence is *relevant*.

[9] Suppose an agent correctly believes it is 12:00 at 12:00. If the agent at 12:00 were to conditionalize on 'it is 1:00', he would have both 'it is 12:00' and 'it is 1:00' in his belief set. Such an agent would be horribly confused. So conditionalization seems to say that if at a later time the agent learns it is 1:00 then the agent should be horribly confused. But this is not the situation at all. See Titelbaum's Sleeping In example (p. 566).

[10] I am grateful to Wolfgang Schwarz for discussion of these issues.

[11] More specifically the *type* of evidence and hypothesis. But omitting this shouldn't lead to any confusion.

	Hypothesis	Evidence	Relevant?
1	Eternal	Eternal (Discovery)	Yes
2	Eternal	Temporal a) Mutation	*No*
		b) Discovery	Yes
3	Temporal	Eternal (Discovery)	Yes
4	Temporal	Temporal a) Mutation	Yes
		b) Discovery	Yes

Temporal evidence can be acquired by Mutation or Discovery; eternal evidence can only be Discovered. This paper is about 2. The main focus will be defending my answer of No at 2a; I will briefly defend my answer of Yes at 2b at the end. It is worth first distinguishing the other cases I am not discussing. 1 is the traditional case where both the evidence and the hypothesis are eternal. Standard examples involving balls in urns or experimental results fit into 1. There is no doubt such evidence is relevant. 3 is concerned with cases where an eternal belief is learnt, and this changes one's degree of belief in a temporal hypothesis. For example, suppose you know a coin flip of Tails results in you being woken at 8am while a coin flip of Heads results in you being woken at 9am. If you have just been woken and are unsure of the time, then learning that the coin landed Heads (eternal) confirms that it is 9am (temporal). So the evidence is relevant. For case 4, suppose the evidence is 'it is 2000' and the hypothesis is 'it is the 21st century'. I take it to be uncontroversial that the evidence is relevant, whether it is acquired by Mutation or Discovery.

The controversial case is 2, and 2a especially. We have seen that we need new rules of belief change to model Mutation, but do we need new rules of belief change governing how *eternal beliefs* are affected by Mutation?

Question Can Belief Mutation produce a shift in credence in eternal beliefs?

The need to address this issue has become apparent due to thought-experiments which appear to show that Mutation *can* produce a shift in the agent's credence in eternal beliefs. If so, the answer to the *Question* is Yes. This answer would mean that credence in an eternal belief can change even when nothing that was uncertain is learnt (because Mutation can occur even when nothing that was uncertain is learnt). If nothing uncertain is learnt, conditionalization permits no belief change. So (eternal) belief change in the absence of something uncertain being learnt amounts to a *violation of conditionalization for eternal beliefs*[12]. We would need new rules to govern such changes. A need for such new rules has been suggested by Elga (2000) and Arntzenius (2003), and new rules have been defended by Halpern (2004), Meacham (2008), Titelbaum (*ibid*.). But despite its popularity, rejecting conditionalization is a radical move. Conditionalization is the heart of our best theory of confirmation (see Earman 1992; Howson and Urbach 1993 for classic Bayesian texts), encodes common sense and is supported by several important arguments (Teller 1976; Williams 1980; Van Fraassen 1999; Greaves and Wallace 2006).

[12] The problems for conditionalization are also problems for Reflection (Van Fraassen 1984). But conditionalization is the more important rule, so I will focus on it. See Schervish et al. (2004) and Weisberg (2007) for relevant discussions.

(Why am I not concerned about giving up conditionalization for Mutation itself then? Because Mutation applies only to beliefs that change in truth-value. The arguments cited all assume that the beliefs don't change in truth-value. For example, someone who bets that it is Monday will not expect the bookie to wait 24 hours and say 'It's Tuesday—you lose!')

I will argue that in fact there is no need for new rules of belief change because the thought experiments that purport to show violations of conditionalization fail. My answer to the *Question* is No; Mutation should have no effect on credence in eternal beliefs (2a). However, I will argue in the final section that *Discovery* of temporal beliefs *can* shift credence in eternal beliefs (2b). I will now consider the thought-experiments purported to show that Mutation can affect credence in eternal beliefs, and argue that they fail.

3 The Prisoner

Imagine you are a prisoner (Arntzenius 2003[13]). The prison guard will flip a fair coin at midnight. If the coin lands Heads he will turn off the light in your cell at midnight. If the coin lands Tails he will leave the light on (Fig. 1).

You are locked in your cell at 6pm. As there is no clock in your cell, you lose track of the time. Imagine it has been a few hours since you were locked in your cell. The light is still on. You think it might be after midnight, but you're not sure. Arntzenius claims that at this point, your degree of belief that the coin landed Tails should go up. I agree. He thinks that this is a new and puzzling way in which temporal beliefs affect eternal beliefs. I disagree. I think that an eternal belief has been Discovered and conditionalization can be applied. Let's look more carefully at how the prisoner's beliefs evolve over time.[14]

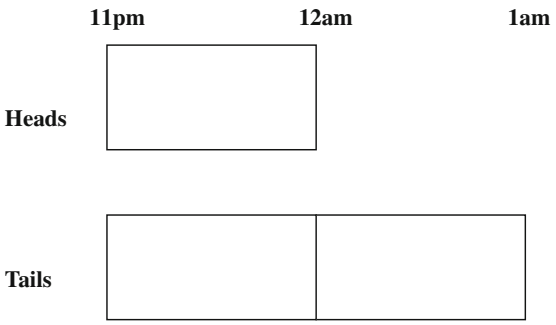


Fig. 1 The *Prisoner* Boxes represent locations where the light is on

[13] Arntzenius gives four other problem cases for conditionalization and reflection in this paper. One is Sleeping Beauty, another is a version of the prisoner, and another is a straight-forward case of memory erasure (Shangri-La), in which the agent violates conditionalization involuntarily. The final case seems to be a combination of the earlier cases.

[14] I want to point out that it is very easy to put yourself in a Prisoner type situation, and I strongly recommend the experience.

First consider a case where there is no light being switched off. What happens to an agent's temporal beliefs as time passes? Two things happen. First of all, they shift forward in time. The belief that it is 6 pm is replaced by the belief that it is 7 pm. This is belief mutation. But when the agent is an imperfect timekeeper, something else happens; the beliefs become more spread out. That is, the agent becomes less certain about exactly what time it is. At 7 pm, the agent might assign an 80% probability to it being within 10 minutes of 7 pm. But by 11 pm, he might only assign a 50% probability to it being within 10 minutes of 11pm (Fig. 2).

Now let's add the extra uncertainty of the coin toss. As well as being uncertain about the time, you are also uncertain about whether the coin landed Heads or Tails. So at 7pm, your probability distribution is spread over various times in two possible worlds, Heads and Tails. Each curve is half the area it was when there was no coin toss to be uncertain about (Fig. 3).

Now let's add the fact that the lights go off at midnight if Heads lands. Consider what happens as the right hand side of the probability distribution edges towards midnight. That is, what happens as you start to think that it may already be later than midnight? If the light remains on, then the possibility that it is later than midnight and Heads will be eliminated. This is because if the coin landed Heads, the light goes off at midnight. If it really is after midnight and the light is still on, then the

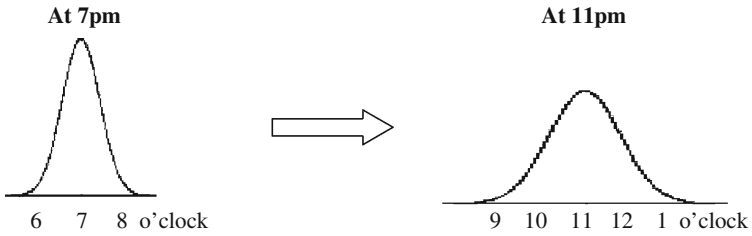


Fig. 2 The passage of time

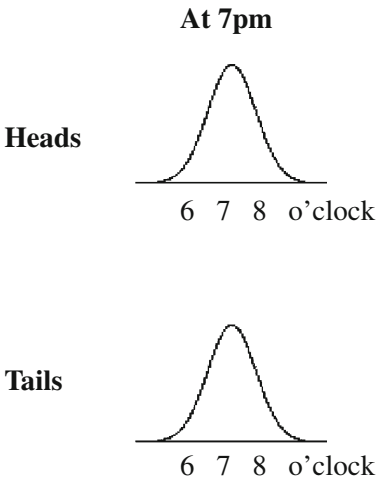


Fig. 3 Was it Heads, and what time is it?

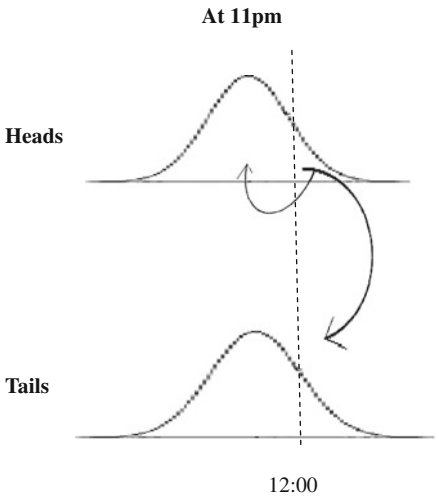


Fig. 4 The shift to Tails. The probability space from the right hand side of the Heads curve is transferred to the uneliminated parts of the curve. The probability of Tails grows to more than 50%

coin must have landed Tails. This means that the probability of Tails must go up[15] (Fig. 4).

Consider your credence at 11:59 pm. You will think it might be after midnight and take the light being on as evidence of Tails. This shift towards Tails is foreseeable at 6pm. As Arntzenius points out, this is an odd situation. It *appears* that nothing that was uncertain at 6 pm has been discovered by 11:59 pm. So there can be no change in credence that is due to conditionalization. If credences change nonetheless, conditionalization is violated and some other rule of belief change is needed.[16] All that appears to have happened is that time has passed, so it looks like a case where mere Belief Mutation is relevant to an eternal belief. So the answer to the Question (can Mutation produce a shift in credence in eternal beliefs?) would be Yes.

But I will argue that an eternal piece of evidence has been Discovered (on the model of Belief Discovery) between 6 pm and 11:59 pm. The shift in credence of Tails is due to conditionalization on this piece of evidence.

4 Diagnosis: What the prisoner learns

I think that the prisoner Discovers a piece of evidence, and this is responsible for the confirmation of Tails. To see what this new belief is, we have to consider why the prisoner changes his degrees of belief in the first place. The shift is caused by the prisoner's new belief that time has passed. Suppose that the prisoner's degree of belief that it is after midnight is 0% at 6pm and greater than 30% at 11pm. Given a very

[15] If the coin landed Heads, the light goes off at midnight. Then you know for certain that it is midnight and the coin landed Heads. Otherwise, the light stays on and your degree of belief in Tails continues to rise. Eventually, you will be confident that it is after midnight and your degree of belief in Tails will approach 1.

[16] I agree with Arntzenius that Reflection is violated however. Weisberg and Seidenfeld et al. both point out that Reflection need not hold when the agent loses track of the time.

plausible introspective awareness, the prisoner knows (at 11 pm) that his credence that it is after midnight is greater than 30%.[17] He also believes the light is still on. Combining these results generates a new belief that has been Discovered.

> *New Discovered Belief* The light is on after my credence that it is after midnight has gone up above 30%.

When the prisoner was first put in the cell at 6pm, he didn't know if the lights would still be on by the time his credence that it is after midnight had gone up above 30%. For all he knew, the light might have been turned off before his credence that is was after midnight got that high. So when the light stays on, he Discovers a new piece of evidence on which he can conditionalize. (This new belief is eternal, but what is essential for my argument is that it is Discovered.)

The probability of the New Discovered Belief being true given Tails is 1; the light will stay on all night if Tails landed. But if Heads landed, the light might have been turned off before his credence that it is after midnight had gone up above 30%. The new evidence, being more likely given Tails, confirms Tails.

> P(The light is on after my credence that it is after midnight has gone up above 30%|Tails) = 1 >
> P(The light is on after my credence that it is after midnight has gone up above 30%| Heads)

This account shows that it is not merely the Mutation of temporal beliefs that causes the shift in the credence in Tails, but the Discovery of a new belief.

One might worry that there is an *inevitable* change in credence. That is, one might worry that the prisoner knows at 6pm that he will learn 'the light is on after my credence that it is after midnight has gone up above 30%'. If he knows he will learn this in the future, he should surely believe it now, and things are looking mysterious again.

But the prisoner cannot be certain that he will acquire any such belief. At 6 pm he might think it possible that the lights will go on before he comes to think there is any chance of it being after midnight. That is, at 6 pm, he may assign a non-zero credence to the eternal proposition that his 11:59 pm credence that it is after midnight will be zero. If so, he will not expect an inevitable shift. (This does not imply that he believes he is a perfect time-keeper, for he may also believe that his credence that it is after midnight could remain zero even after midnight.)

I conclude that the Prisoner does not show that Belief Mutation has any effect on credence in eternal beliefs. Arntzenius's example appears to show that the Mutation of temporal beliefs can affect eternal beliefs in unexpected ways. But I have argued that the prisoner Discovers a new eternal belief that he didn't know at 6pm. The prisoner updates by conditionalizing on this new belief. No new norm of eternal belief change is required. I'll now argue that in a second case, the argument that Belief Mutation leads to a violation of conditionalization is inconclusive.

[17] Note that the Prisoner doesn't need, and I haven't assumed, perfect transparency of his own credences. Some vague idea that his credence that it is after midnight is greater than it was earlier is sufficient.

5 Sleeping Beauty

It is Sunday night. Sleeping Beauty is about to be drugged and put to sleep. She will be woken briefly on Monday. Then she will be put back to sleep and her memory of being awoken will be erased. She might be awoken on Tuesday. Whether or not she is depends on the result of the toss of a fair coin. If it lands Heads, she will not be woken. She will sleep straight through to Wednesday, and the experiment will be over. If it lands Tails, she will be awoken on Tuesday. The Monday and Tuesday awakenings will be indistinguishable. Sleeping Beauty knows the setup of the experiment and is a paragon of probabilistic rationality (Fig. 5).

There are three centred worlds where Beauty could be:

H1 = Monday and Heads
T1 = Monday and Tails
T2 = Tuesday and Tails

Obviously Beauty's credence in Tails on Sunday should match the objective chance of 1/2. This reasoning employs a *probability co-orindation principle* (PCP) that connects subjective credences with the objective chances.[18] But the question is: when she is woken, what credence should she have that the coin landed Tails?

Some say that her credence in Heads should stay at 1/2. Call these *Halfers.*
Some say that her credence in Heads should fall to 1/3. Call these *Thirders.*

What exactly has changed between Sunday and waking up on Monday and Tuesday? Beauty has acquired the temporal belief that [it is now Monday or Tuesday]. This type of belief acquisition is not the type of belief acquisition modelled by conditionalization. Conditionalization models cases in which evidence that was initially uncertain is learnt. But Sleeping Beauty has not learnt anything about which she was

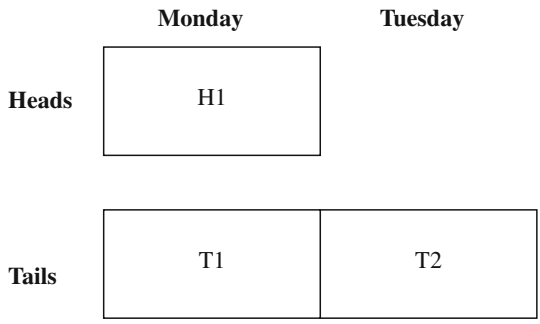


Fig. 5 The *boxes* represent days when Beauty is awake

[18] The terminology is from Strevens (1995). The most well known PCP is that of Lewis (1980, 1994). But I'm not using his Principal Principle because I do not wish to give the impression that my arguments, or those of Elga (2000), depend on the details of Lewis's account. Elga doesn't even mention Lewis in this context, and Lewis himself coyly speaks of 'a well-known principle which says that credences about future chance events should equal the known chances' (2001, p. 175).

uncertain; this is not Discovery. Instead, she has acquired a new belief because the content of that belief (it is now Monday or Tuesday) has changed from being false to true. This is Belief Mutation.

Thirders claim that such Belief Mutation should shift Beauty's credence in the eternal belief that Heads landed. This would again be a violation of conditionalization and indicate that a new rule of belief update for eternal beliefs is needed. I will defend conditionalization. I cannot address every argument, which include Arntzenius (2002), Arntzenius (2003), Elga (2000), Dorr (2002), Draper and Pust (2008), Hitchcock (2004), Horgan (2004), Horgan (2007), Monton (2002), Titelbaum (*ibid.*) and Weintraub (2004). All these arguments are controversial and most have been challenged in print (see Bradley 2003, 2010, forthcoming; Bradley and Leitgeb 2006; Briggs ms; Lewis 2001; Jenkins 2005; Schwarz ms; White 2006 for a range of responses). But probably the most important thirder argument has not yet been adequately challenged. Both in print and in conversation, Elga's (2000) original argument seems most influential in persuading people of the thirder position. I will argue in this paper that Elga's argument is unpersuasive.

It is important to understand that I am not aiming to refute the thirder position—I have no knock-down arguments that being a thirder is irrational. I am merely trying to show that those persuaded by Elga's arguments should not have been.

6 Elga's argument

Elga's argument requires modifying Sleeping Beauty a little. But the modifications are harmless and the result is very interesting. The first modification is based on the fact that it doesn't matter when the coin is tossed. The experimenters could toss the coin on Sunday night, and then wake Beauty either once or twice. Or they could wait until Monday night, toss the coin, and wake her on Tuesday only if it lands Tails. So let's assume that they do the latter. The coin is tossed on Monday night, and Beauty is only woken on Tuesday if it lands Tails. Assume Beauty knows this.

The second modification is that after waking on each day Beauty is told what day it is. What should Beauty think after she is told that today is Monday? She knows that a coin is going to be tossed *tonight*. She also knows that none of her memories have been erased. If there are to be any cognitive mishaps, they lie in the future. Given this situation, Elga argues that Beauty's credence should match the objective chances, so she should assign a probability of 1/2 to the coin landing heads.

The argument for being a thirder then runs as follows. Let P be her subjective probabilities just after she is woken on Monday. Let P+ be her probabilities after she is told it's Monday. Let P− be her probabilities on Sunday night. Elga claims that $P+(H1) = 1/2$ from the PCP. Then he argues backwards to the situation before Beauty found out it was Monday. After learning it was Monday, her credence that the coin will land Heads ought to be the same as the conditional credence $P(H1 \mid H1 \text{ or } T1)$.[19] So $P(H1 \mid H1 \text{ or } T1) = 1/2$, and hence $P(H1) = P(T1)$. Elga then applies his Restricted

[19] This is a case where a temporal belief is learnt from a position of uncertainty i.e. Discovery. But the shift we are concerned with here, from Sunday to [Monday or Tuesday], is Mutation.

Principle of Indifference (defended in his 2004), which says that agents in subjectively indistinguishable states within the same possible world should have equal credences.[20] The agents at T1 and T2 satisfy these conditions, so P(T1) = P(T2). So we have P(H1) = P(T1) = P(T2). As these are mutually exclusive and exhaustive, P(H1) = 1/3. So runs Elga's argument for 1/3.

I think this argument has been refuted by Lewis (2001). But Lewis was not as clear as usual, and his argument has not been widely accepted or even discussed.[21] I will defend and expand upon Lewis's argument. Due to the brevity of Lewis's argument, it may be extravagant to attribute my argument to him. So I will more cautiously claim that the following is an argument that I have been led to by Lewis's paper. I should emphasize that I am not aiming to give a positive argument for the halfer position, but merely to undermine Elga's argument for the thirder position.

Lewis rejects Elga's premise that P+(Heads) = 1/2. Elga's justification for the premise is a PCP. But there are some situations in which credence should diverge from objective chance. For example, suppose we have a crystal ball known to be reliable that predicts that this flip of a fair coin will land Tails. What should we believe about the outcome of the coin flip? Should we follow the crystal ball, or should we stick with the objective chance? We should follow the crystal ball. Evidence about the future can trump the objective chance. Call evidence which can justify an agent in having credences that diverge from the chances *inadmissible evidence*. The terminology is from Lewis (1980, 1994); I am not endorsing Lewis's version of the PCP though, just his terminology.

7 Beauty's inadmissible evidence

Does Sleeping Beauty have inadmissible evidence (relative to the coin toss[22])? I say yes. The paradigm sources of inadmissible evidence are crystal balls, oracles, and suchlike. Sleeping Beauty has nothing as obviously inadmissible as this. But I will argue that Sleeping Beauty has inadmissible evidence when she is told that today is Monday. The way to see this is to consider the alternative evidence she might have received. From the state of being awake in the Sleeping Beauty setup, there are two pieces of evidence she might have found. She might have been told it's Monday or she might have been told it's Tuesday. So the evidence space is the following:

{Today is Monday, Today is Tuesday}

My argument that Beauty has inadmissible evidence when she learns it is Monday will proceed in two steps:

1. 'Today is Tuesday' is inadmissible.

[20] Equal credences of what? Specifically, equal credences of whether each of them is one of the agents rather than the other. It follows that they must have equal credence in everything, so I have left the main text unqualified.

[21] Dieks (2007) is an exception.

[22] All inadmissible evidence is relative to an event, and possibly to other variables too. I will leave these variables implicit in future.

2. If an agent with only admissible evidence has two possible pieces of evidence in her evidence space and one piece of evidence is inadmissible, then the other is inadmissible.

Argument for 1: Suppose Beauty learns that today is Tuesday. Should her degree of belief in Heads be 50%? No. Her degree of belief in Heads should be zero, because if Heads landed, she would sleep through Tuesday. As this evidence justifies Beauty not setting her degree of belief to match the objective chances, it is inadmissible evidence.[23]

Argument for 2: If there are two possible pieces of evidence, E1 and E2, an agent's prior degree of belief in hypothesis H must be a weighted average of P(H|E1) and P(H|E2). Assume that E1 is inadmissible. Then P(H|E1) need not be equal to the objective chance of H. Perhaps P(H|E1) is less than P(H). Then P(H|E2) must be more than P(H) (otherwise P(H) won't be the weighted average). As the agent initially had only admissible evidence, P(H) should equal the objective chance of H. So P(H|E2) should be greater than the objective chance of H. Which means E2 is inadmissible.

It follows from 1 and 2 that Beauty has inadmissible evidence when she is told that today is Monday. This is why she is not bound by the PCP, so Elga's premise that $P+ = 1/2$ is unsupported and his argument is inconclusive. Intuitively, as 'it is Tuesday' would confirm Tails (absolutely), 'it is Monday' confirms Heads. The halfer can claim that credence in Heads after learning it's Monday is 2/3 (the result of conditionalizing on 'it is Monday' from a prior probability of Heads of 1/2).

Thirders might object that Beauty has inadmissible evidence when she wakes up on Monday, so she does not satisfy the antecedent of 1 (Dieks *ibid.*). Thirders might continue that merely being woken gives Beauty evidence that favours Tails, and is therefore inadmissible evidence.

But no reason has been offered for the halfer to accept this—and this should be the conclusion of the argument, not a premise. Halfers do not think Beauty gets inadmissible evidence on waking; thirders need to offer an argument that she does. They cannot simply assume that she gets inadmissible evidence, for that is the very point that the argument is supposed to show. I'm not saying that no argument could possibly be produced of course; again, I'm not offering a knock-down argument that being a thirder is irrational. I'm just arguing that those persuaded by Elga's argument should not have been. The PCP does not bare the weight that Elga's argument places upon it. Keep in mind that Elga is arguing that we should do something radical i.e. substantially revise our best theory of confirmation, so we should require a strong argument to do so.

A different objection to my argument is to point out that it is only valid if the evidence doesn't alter the agent's beliefs regarding the objective chance. This is true—'Today is Monday' and 'Today is Tuesday' mustn't change Beauty's beliefs about what the

[23] One might object that if Beauty had learnt it was Tuesday and evidence must be true, then it would be Tuesday. If on Tuesday the objective chance of Heads is zero, then Beauty's credence would match the objective chances, so her evidence would be admissible. This objection could be blocked by denying that evidence in this context need be true or denying time-dependent chances (Hoefer 2007). I would endorse both moves. I am grateful to Wolfgang Schwarz for discussion here.

chances are. This is satisfied—it is stipulated that Beauty *knows* the coin is fair. So she should only doubt that it is fair if she receives evidence connected to it being biased, which she has not.

A third objection is to point out that the argument is only valid if the pieces of evidence are mutually exclusive and exhaustive.[24] This can be granted as it is obviously satisfied by 'Today is Monday' and 'Today is Tuesday'. We can conclude that 'today is Monday' is inadmissible evidence.

One might still worry that there must be *something* wrong with saying that Beauty's credence in a future coin flip should diverge from the objective chances. This should only happen if she has evidence about the future.

I respond that Sleeping Beauty does have evidence about the future. Where did it come from? There are no prophets or crystal balls around. It came from the evidence space. One of the possible pieces of evidence *is* about the future—the possible evidence that today is Tuesday. This, I think, is what Lewis meant when he said that Beauty has evidence about the future, 'namely that she is not now in it'[25] (p. 175). It should not be surprising that finding out the time gives you evidence about the future. How could it not? Finding out the time involves finding out what time is present and what times are future.

There are of course many other arguments for the thirder position that I cannot address here, but I hope to have shown that the most influential argument for the thirder position does not succeed. Thus the thesis that conditionalization is the only rule of update for eternal beliefs survives. Before drawing some general morals, I want to bring out the connection between Sleeping Beauty and the Prisoner.

8 (Dis)Analogy

There is a striking similarity between Sleeping Beauty and the Prisoner. Both involve two possible worlds, and an extra temporal possibility in one of them (Figs. 6 and 7).

Yet I have argued that the Prisoner gets confirmation of Tails but Beauty doesn't. What's the difference? One crucial difference is that the Prisoner has memories that give him some indication of the passage of time. So he has some relevant evidence about whether he is in the earlier or later stage of the experiment (Monday or Tuesday for Beauty, before midnight and after midnight for the Prisoner). His memories, combined with the fact that the light is still on, result in his Discovering the new belief 'the light is on after my credence that it is after midnight has gone up above 30%'. Beauty has no memories, so she has Discovered no equivalent belief which she can conditionalize on. So the argument for why the Prisoner should change his credences cannot be applied to Beauty.[26] Let's now put these results in a broader context.

[24] I am grateful to an anonymous *Synthese* referee for pointing this out.

[25] As Elliott Sober pointed out to me, this is an unfortunate phrase—one always knows that one is not in the future. This is plausibly true even for those with time-machines.

[26] Another disanalogy is that Sleeping Beauty has zero chance of observing Tuesday if Heads, whereas the Prisoner has a non-zero chance of observing between 12am and 1am if Heads. It may be dark, but he will be conscious. So if the Prisoner, like Beauty, has no memories, his observation that the light is on will *still* confirm Tails. The selection procedures by which they make their observations are different. I discuss the connection between selection procedures and self-location in my (forthcoming).

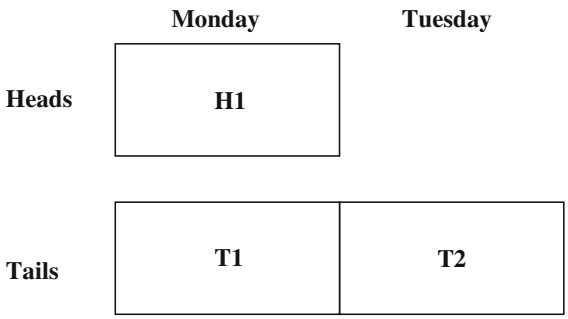


Fig. 6 The boxes represent days when Beauty is awake

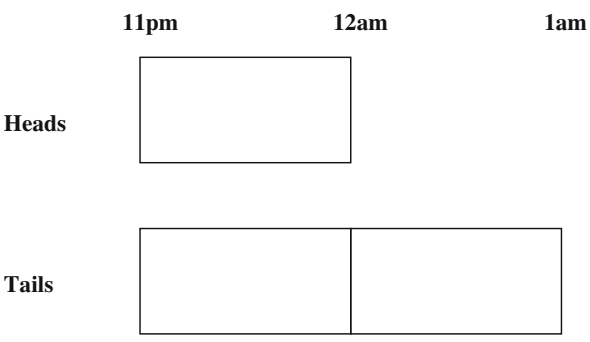


Fig. 7 The boxes represent centred worlds locations where the light is on

9 Comparison

Let's back up. I have distinguished belief change in virtue of learning the truth of something previously uncertain (Discovery) from belief change due to the content of the belief changing in truth-value (Mutation). Can temporal belief change of either type shift credence in eternal beliefs? If they can, we'll say that they are relevant to eternal beliefs. Those who have addressed the question so far have dealt with Discovery and Mutation together. Titelbaum holds that both types are relevant to eternal belief[27]; Halpern (2004); Bostrom (2007); Meacham (2008)[28] argue that neither type is. But Discovery and Mutation are very different and deserve separate treatment. I hold that Discovery is relevant to eternal beliefs, but Mutation is not. We can zoom in and expand Table 1 to compare my position to others.

[27] I focus on Titelbaum because among thirders he is most explicit that he is defending general norms of belief change. He attacks the strong thesis that 'it is never rational for an agent who learns only self-locating [temporal] information to respond by altering a non-self-locating [eternal] degree of belief' (p. 556). But this is only in conflict with Halpern (*ibid.*), Bostrom (*ibid.*) and Meacham (*ibid.*). The more popular halfer view accepts that Discovery is relevant e.g. Lewis (2001), as Titelbaum (p. 585) notes. I discuss Titelbaum's analysis in Bradley (forthcoming).

[28] Meacham is the most explicit that he is defending general norms of belief change, but I think it natural to read Halpern and Bostrom as committed to the same conclusion. There are differences in their theories that I will largely ignore in the discussion that follows. I think they each face at least two of the problems I raise.

	Titelbaum	Bradley	Halpern, Bostrom, Meacham
2a) Mutation is relevant	Yes	No	No
2b) Discovery is relevant	Yes	Yes	No

My arguments above attempted to show that the arguments in favour of the relevance of Mutation do not succeed. I have not given arguments against the relevance of Mutation in this paper; the argument is that it commits us to a violation of conditionalization. Indeed the initial arguments for the relevance of Mutation were put forward with an explicit acknowledgement of this (see Arntzenius and Elga in particular). Although this position is popular, giving up a principle as defensible and reasonable as conditionalization is a heavy cost.

One might object that I've already admitted we must give up conditionalization and reflection for temporal beliefs, so why not admit the relevance of Mutation? Because the issue at stake is whether we should give up conditionalization for *eternal* beliefs. It is not controversial that beliefs such as 'today is Monday' must change in ways that don't conform to conditionalization. The issue is whether eternal beliefs such as 'the coin landed Heads' can change in ways that violate conditionalization. I have argued that the arguments in favour of such a violation do not succeed.

I have not said anything to undermine arguments from the other side (i.e. Halpern (*ibid.*), Bostrom (*ibid.*) and Meacham (*ibid.*)) who seem to claim that temporal beliefs can *never* have any effect on eternal beliefs. That is, that Discovery of temporal beliefs is not relevant either. So I will now offer a brief rear guard action before concluding.[29]

The main problem with the position that temporal beliefs are never relevant is the implausibility of its commitments. These result from denying conditionalization, and, moreover, from denying it in cases where we have strong intuitions that it should hold (in contrast to cases like Sleeping Beauty and the Prisoner in which we do not have clear intuitions). If Discovery of temporal beliefs isn't relevant, then when Beauty learns it is Monday, her credence in Heads should not shift, even though the conditional probability of Heads given Monday is different from the unconditional probability. So proponents are committed to the following two constraints:

Constraint a $P(\text{Heads}|\text{Monday}) = 2/3$
Constraint b $P_{\text{Monday}}(\text{Heads}) = 1/2$

This is bad, but things get worse. Consider a variation in which Beauty is woken 10 times given either Heads or Tails. If Heads lands she sees a red light on 9 days and a blue light on 1; if Tails lands she sees a blue light on 9 days and a red light on 1. As before, her memory is erased between each waking. So for any given day, there is a 90% chance of seeing a red light given Heads and a 10% chance of seeing a red light given Tails. Intuitively, seeing a red light should confirm Heads. But she only learns the temporal 'There is a red light *today*' when she sees the light. She doesn't learn any eternal evidence on seeing the light, such as 'There exists some day on which the red light is seen'—she knew that all along. So if temporal evidence is never relevant, seeing the light fails to confirm Heads.

[29] I am grateful to Matt Kotzen for an exchange on the issues that follow.

A second problem is that it is too easy to turn Discovery of a temporal belief into Discovery of an eternal belief. We can do this by supposing that the original Sleeping Beauty has some unique experience on each day. For example, she might see what the clouds look like while knowing that the clouds will not be identical on both days (Dorr ms) or she might wear different coloured pyjamas each day (Monton and Kierland 2005) or she might observe a differently coloured piece of paper on each day (Titelbaum). In each case, Beauty would acquire eternal evidence e.g. there exists some day on which the red paper is seen, but it is implausible to think that different rules of belief change apply to the modified cases with the unique experience than to cases without.[30]

A third problem is that the main theoretical justification for the irrelevance of Discovery of temporal beliefs, offered by both Bostrom and Halpern, fails. They both argue that there is a familiar difference between 'E' and 'I learn E', and that this explains how *a* and *b* can both be accepted. This would allow them to keep conditionalization after all. Halpern cites the Monty Hall problem as a case in which this distinction makes a difference. But the natural thing to do in response is to include the evidence about how E is learnt in the updating:

$$\begin{array}{ll} \textit{Constraint } c & \mathrm{P}(\mathrm{Heads}|\mathrm{I\ learn\ it's\ Monday\ today}) = 2/3 \\ \textit{Constraint } d & \mathrm{P}_{\mathrm{I\ learn\ it's\ Monday\ today}}(\mathrm{Heads}) = 1/2 \end{array}$$

The conflict simply re-emerges one level up.

Even if Bostrom and Halpern can block this response, I don't think the distinction between 'E' and 'I learn E' can do the work they want it to in this case. The reason is that the distinction only makes a difference if it is epistemically possible that either E can be true without learning E, or learning E can be true without E.[31] Otherwise, E and learning E are equivalent. Yet Sleeping Beauty discovers it is Monday iff it is Monday. So there isn't the required difference between E and learning E that is needed to accept constraints *a* and *b*.

Bostrom and Halpern might offer a second reply: that I have missed their point, and that the agent should not use his conditional probability of H given E when he later learns E. Bostrom suggests that doing so fails to take into account that the agent learns not just E, but also that he is at a time when he has learnt E.

But this move would amount to a whole-scale rejection of conditionalization in any context. Any sufficiently reflective agent can turn 'E' into 'I am at a time when I have learnt E' just by reflecting on what she has learnt. This can happen in any context, *even if E is eternal*. So any new norms regarding 'I have learnt E' will be global, and infect the whole of confirmation theory.[32]

I conclude that Discovery of temporal beliefs can be relevant to eternal beliefs.

[30] One might object that cases with unique experiences generate an argument via conditionalization for thirding e.g. Titelbaum. I argue in Bradley (forthcoming) that this argument fails.

[31] In the Monty Hall problem it is epistemically possible, for each door, that it is empty without you discovering that it is empty.

[32] I discuss the distinction between 'E' and 'I learn E' in Bradley (2010).

10 Conclusion

I have argued that there are importantly different ways in which temporal beliefs can be acquired. Sometimes they are acquired by Discovery, in the same way as has been traditionally discussed by Bayesians. I have argued that such learning can shift credence in eternal beliefs. But temporal beliefs, unlike eternal beliefs, can also be acquired by Mutation. I have argued that Mutation cannot change credence in eternal beliefs, so we need no new norms of belief change for eternal beliefs. The costs of doing so are high, and the arguments so far offered are unconvincing.

Acknowledgements I am grateful to Frank Arntzenius, Paul Bartha, Nick Bostrom, Kenny Easwaran, Andy Egan, Adam Elga, Justin Fisher, Branden Fitelson, Hilary Greaves, Alan Hájek, Matt Kotzen, Chris Meacham, John Perry, Teddy Seidenfeld, Elliott Sober, Mike Titelbaum, Jonathan Weisberg and an audience at the ANU for helpful discussion and comments on this material.

References

Arntzenius, F. (2002). Reflections on Sleeping Beauty. *Analysis, 62*(273), 53–62.
Arntzenius, F. (2003). Some problems for conditionalization and reflection. *Journal of Philosophy, 100*, 356–370.
Bradley, D. J. (2003). Sleeping Beauty: A note on Dorr's argument for 1/3. *Analysis, 63*, 266–268.
Bradley, D. J., Leitgeb, H. (2006). When betting odds and credences come apart: More worries for Dutch book arguments, *Analysis, 66*(2), 119–127.
Bradley, D. J. (2010) "Conditionalization and beliefs *De Se*" *Dialectica.*
Bradley, D. J. (forthcoming). Confirmation in a branching world: The Everett interpretation and Sleeping Beauty. *British Journal for the Philosophy of Science*.
Briggs (ms). Putting a value on Beauty.
Bostrom (2007). Sleeping beauty and self-location: A hybrid model. *Synthese, 157*(1), 59–78.
Dorr, C. (2002). Sleeping Beauty: In defence of Elga. *Analysis, 62*, 292–296.
Dieks, D. (2007). Reasoning about the future. *Synthese, 156*, 427–439.
Draper, K., & Pust, J. (2008). Diachronic Dutch books and Sleeping Beauty. *Synthese, 164*(2), 282–287.
Earman, J. (1992). *Bayes or bust*. Cambridge, MA: MIT Press.
Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis, 60*, 143–147.
Elga, A. (2004). Defeating Dr. Evil with self-locating belief. *Philosophy and Phenomenological Research, 69*(2).
Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind, 115*(459), 607–632.
Halpern, J. (2004). Sleeping Beauty reconsidered: Conditioning and reflection in asynchronous systems. In Proceedings of the Twentieth conference on uncertainty in AI, pp. 226–234.
Hitchcock, C. (2004). Beauty and the bets. *Synthese, 139*(3), 405–420.
Hoefer, C. (2007). The third way on objective probability: A sceptic's guide to objective chance. *Mind, 116*(463), 549–596.
Horgan, T. (2004). Sleeping Beauty awakened: New odds at the dawn of the new day. *Analysis, 64*, 10–21.
Horgan, T. (2007). Synchronic Bayesian updating and the generalized Sleeping Beauty problem. *Analysis, 67*(293), 50–59.
Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago: Open Court.
Jenkins, C. (2005). Sleeping Beauty: A wake-up call. *Philosophia Mathematica, 13*(2), 194–201.
Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–566). Oxford: Oxford University Press.
Lewis, D. (1979). Attitudes *De dicto* and *De se*'. In D. Lewis (Ed.), *Philosophical Papers* (Vol. 1). Oxford University Press (1983).

Lewis, D. (1980). A subjectivist's guide to objective chance. *Studies in Inductive Logic and Probability* (Vol. 2). Berkeley, CA, USA: University of California Press.
Lewis, D. (1994). Chance and credence: Humean supervenience debugged. *Mind, 103*, 473–490.
Lewis, D. (2001). Sleeping Beauty: Reply to Elga. *Analysis, 61*, 171–176.
Meacham, C. (2008). Sleeping Beauty and the dynamics of *De se* belief. *Philosophical Studies, 138*(2), 245–269.
Monton, B. (2002). Sleeping Beauty and the forgetful Bayesian. *Analysis, 62*, 47–53.
Monton, B., & Kierland, B. (2005). Minimizing inaccuracy for self-locating beliefs. *Philosophy and Phenomenological Research, 70*(2), 384–395.
Perry, J. (1979). The problem of the essential indexical. *Nous, 13*, 3–21.
Schaffer, J. (ms) The Schmentencite way out: Towards an index-free semantics.
Schwarz (ms) Changing minds in a changing world.
Schervish, M. J., Seidenfeld, T., & Kadane, K. B. (2004). Stopping to reflect. *Journal of Philosophy, 101*(6), 315–322.
Strevens, M. (1995). A closer look at the 'New Principle'. *British Journal for the Philosophy of Science, 46*(4), 545–561.
Teller, T. (1976). Conditionalization, observation, and change of preference. In William Harper & C. A. Hooker, *Foundations of probability theory, statistical inference, and statistical theories of science*. Dordrecht: D. Reidel.
Titelbaum, M. (2008). The relevance of self-locating beliefs. *Philosophical Review, 117*(4), 555–606.
Van Fraassen, B. C. (1984). Belief and the will. *Journal of Philosophy, 81*, 235–256.
Van Fraassen, B. C. (1999). A new argument for conditionalization. *Topoi, 18*, 93–96.
Weintraub, R. (2004). Sleeping Beauty: A simple solution. *Analysis, 64*, 8–10.
Weisberg, J. (2007). Conditionalization, reflection, and self-knowledge. *Philosophical Studies, 135*, 179–197.
White, R. (2006). The generalized Sleeping Beauty problem: a challenge for thirders. *Analysis, 66*, 114–119.
Williams, P. (1980). Bayesian conditionalisation and the principle of minimum information. *The British Journal for the Philosophy of Science, 31*(2), 131–144.