Ross on Sleeping Beauty

Brian Weatherson

2013

In two excellent recent papers, Jacob Ross has argued that the standard arguments for the 'thirder' answer to the Sleeping Beauty puzzle lead to violations of countable additivity. The problem is that most arguments for that answer generalise in awkward ways when he looks at the whole class of what he calls Sleeping Beauty problems. In this note I develop a new argument for the thirder answer that doesn't generalise in this way.

In two excellent recent papers, Jacob Ross (2010, 2012) has argued that the standard arguments for the ⅓ answer to the Sleeping Beauty puzzle lead to violations of countable additivity. The problem is that most arguments for the ⅓ answer generalise in awkward ways when he look at the whole class of what he calls *Sleeping Beauty problems*.

> Let us define a *Sleeping Beauty problem* as any problem in which a fully rational agent, Beauty, will undergo one or more mutually indistinguishable awakenings and in which the number of awakenings she will undergo is determined by the outcome of a random process. Let S be a partition of alternative hypotheses concerning the outcome of this process. Beauty knows the objective chance of each hypothesis in S, and she also knows how many times she will awaken conditional on each of these hypotheses, but she has no other relevant information. The problem is to determine how her credence should be divided among the hypotheses in S when she first awakens. The original Sleeping Beauty problem is the instance of this class of problems in which the random process that determines how many times Beauty awakens is a fair coin toss and in which Beauty awakens once given Heads and twice given Tails. (Ross 2010, 413–14)

The 'Generalised Thirder Principle' says

> In any Sleeping Beauty problem, defined by a partition S, upon first awakening, Beauty's credence in any given hypothesis in S should be proportional to the product of its objective chance and the number of times Beauty awakens if this hypothesis is true. (Ross 2010, 414)

Ross shows that this Principle leads to violations of countable additivity, and argues (convincingly to my mind) that this is a serious problem for the Principle. In Ross (2010) he argues that many arguments for the ⅓ answer imply the Generalised Thirder Principle, and in Ross (2012) he argues that the same is true of the argument for the ⅓ answer in Weatherson (2011). He suggests that we have some inductive reason to think that *all* arguments for the ⅓ answer will imply the Generalised Thirder Principle. In reply I want to make three points.

First, Ross offers a much more careful presentation and analysis of the argument for the ⅓ answer than appears in Weatherson (2011), and that analysis does really suggest that the argument overgenerates. But second, the careful setting out suggests that the argument fails by its own standards; it commits a fallacy of equivocation. So I don't think it's as much inductive evidence for Ross's general conclusion about all arguments for the ⅓ answer as he suggests. And third, the inductive conclusion Ross draws is not true. There are arguments for the ⅓ answer that don't imply the Generalised Thirder principle. Below I outline two ways to fix the argument in Weatherson (2011) so that it becomes such an argument.

It will be worthwhile having a minimal example of where the Generalised Thirder principle goes beyond the arguments for the ⅓ answer, so consider the following minimal version.

> **Three-Day Sleeping Beauty**
>
> A coin will be flipped. If it comes up Heads, Beauty will wake up Monday, and go back to sleep until Thursday. If it comes up Tails, Beauty will wake Monday, then have her memories of that waking erased, then wake again Tuesday, and have her memories of that waking erased, then wake up Wednesday, and go back to sleep until Thursday. The different possible wakings will be indistinguishable.[1] When she wakes on Monday, what should her credence be that the coin landed Heads?

The Generalised Thirder says that it should be ¼. One of our aims here will be to come up with an argument for the ⅓ answer in the original puzzle that doesn't imply the answer to the Three-Day Sleeping Beauty puzzle is ¼.

Returning to the original puzzle, Let Cr_1 be Beauty's credence function when she wakes on Monday. Let M be the proposition Beauty would express on waking with "Today is Monday". And let H be the proposition that the coin landed heads. The ⅓ answer entails that the following claims are all true.

1. $Cr_1(M \wedge H) = Cr_1(M \wedge \neg H)$
2. $Cr_1(\neg M \wedge \neg H) = Cr_1(\neg M \wedge \neg H)$
3. $Cr_1(M \wedge \neg H) = Cr_1(\neg M \wedge \neg H)$

[1] I have some concerns about what 'indistinguishable' means in this context. We'll come back to that issue a lot in what follows.

The usual argument for the ⅓ answer argues for (1) and (2), and then derives (3). And the arguments for (2) typically generalise into arguments for the Generalised Thirder Principle, and hence to violations of countable additivity.

To get a bit better feel for what's going on with these principles, consider one interesting sub-class of Ross's large category of Sleeping Beauty problems. These are problems where there are n chance hypotheses, and for any i such that $1 \leqslant i \leqslant n$, Beauty is woken i times if chance hypothesis h_i is true. When she wakes up, we can let the proposition p_{ji}, where $1 \leqslant j \leqslant i \leqslant n$, be the proposition that this is the i'th day, and chance hypothesis h_j is true. The following table, where the chance hypotheses are on the rows, and days are on the columns, represents the possibilities as they strike Beauty.

	Day 1	Day 2	...	Day n
h_1	p_{11}			
h_2	p_{12}	p_{22}		
...				
h_n	p_{1n}	p_{2n}	...	p_{nn}

The usual argument for the ⅓ answer to Sleeping Beauty includes a 'vertical' argument and a 'horizontal' argument. The 'vertical' argument attempts to show that $Cr_1(p_{1i} \mid p_{11} \vee ... \vee p_{\sim 1n}) = \mathrm{Ch}(h_i)$. The 'horizontal' argument attempts to show that $Cr_1(p_{ji} \mid p_{11} \vee ... \vee p_{\sim 1i}) = 1/i$. Or, at least, it attempts to show that those claims are true for the special case where $n = 2$. but as Ross shows, the arguments offered seem to work in the $n = 2$ case iff they work in the general case. And these vertical and horizontal arguments together do imply the Generalised Thirder Principle.

The argument in Weatherson (2011) took a different tack. It argued for (1) and (3), and derived (2) Cr. In terms of the table above, the idea was to replace the 'horizontal' argument, and indeed to reject its conclusion in the general case, with a 'diagonal' argument, which showed that $Cr_1(p_{ii} \mid p_{11} \vee ... \vee p_{nn}) = \mathrm{Ch}(h_i)$. If the vertical and diagonal arguments worked, and didn't overgeneralise, then they wouldn't entail a solution for cases like Three-Day Sleeping Beauty, or about any case from the above class where $n > 2$.

If we drop the restriction to Sleeping Beauty problems where Beauty is woken i times if chance hypothesis $h_{\sim i}$ is true, and return to Ross's more general class, the arguments in Weatherson (2011) were meant to prove the following two claims, and not a lot more.

First Day In any Sleeping Beauty problem where Beauty wakes at least one time in every chance hypothesis, and exactly one time in at least one of them, when she wakes the first time, her conditional credences in each hypothesis in S, conditional on this actually being the first waking, equals the objective chance of each such hypothesis.

Last Day In any Sleeping Beauty problem where Beauty wakes at least one time in every chance hypothesis,, when she wakes the first time, her conditional credences in each hypothesis in S, conditional on this actually being the last waking, equals the objective chance of each such hypothesis.

Since **First Day** entails (1), and **Last Day** entails (3), these principles entail the ⅓ answer. But they don't settle what to say about the Three-Day Sleeping Beauty example. If Cr_3 is Beauty's credences when she wakes for the first time in that example, these principles are consistent with Beauty having the following credal distribution.

$^*\text{Cr}_3$*(Today is Monday and Heads)* = 1/5
Cr_3*(Today is Monday and Tails)* = 1/5
Cr_3*(Today is Tuesday and Tails)* = 2/5
Cr^*_3(Today is Wednesday and Tails) = 1/5

But those credences are incompatible with the Generalised Thirder Principle, so **First Day** and **Last Day** do not entail that principle.

Ross (2012) argues that this isn't right, and that the motivation for **First Day** offered in Weatherson (2011) in fact does lead to the Generalised Thirder Principle on its own. I think that's true in a sense; the argument provides just as much support for the Generalised Thirder Principle as it does for **First Day**. But that's because it is a bad argument, and doesn't support **First Day**. We'll see why that's true shortly. But since it is true, a new argument is needed for (1) , one that supports **First Day**, but not the Generalised Thirder principle.

Most discussions of Sleeping Beauty assume that the contents of propositional attitudes are sets of centered worlds. Following Stalnaker (2008), the argument in Weatherson (2011) tried to get by with propositions simply being sets of worlds. The key idea was that the worlds themselves would be fine-grained enough that thoughts like *Hesperus is Phosophorus*, or *Today is Monday* would be contingently true if true at all. Very roughly, we associate singular terms with something like Fregean senses.[2] When Beauty wonders whether today is Monday, she isn't wondering about whether an instance of the law of identity is true. She has no more interest in that law than does the first gentleman of Europe. She is wondering, in effect, whether two senses, TODAY and MONDAY, have the same referent.

This way of looking at things implies that credal dynamics need to be complicated. Sometimes our credences change because we acquire more information, and we react accordingly. But sometimes our credences change because we acquire new senses, and we can think new thoughts. That will become crucial in what follows.

There is a quick argument for **Last Day**. Consider what happens after Beauty wakes up at the end of the puzzle, and is told that the game is over. (That is, on Wednesday in the original puzzle, or on Thursday in the three-day variant.) Plausibly, her credence in H should be back to ½, or at least it is hard to see a good argument why it should be anything else.[3] She can also think back to her last waking, and think to herself *If*

[2] In Weatherson (2011) these are described as haeceitties, but this is misleading at best. It is crucial that 'Hesperus' and 'Phosphorus' have different associations, but intuitively they have the same haecceity. Thinking of the associations as being with something like senses is better.

[3] One might be tempted by a seemingly stronger argument. For instance, one could argue that on Wednesday, Beauty knows that the chance of H was ½, and she has no inadmissible evidence, so by the Principal Principle her credence in H should be ½. But it isn't clear that she has no inadmissible evidence; perhaps the evidence she would express by thinking back to her last waking and saying *That waking happened* is

H, that was on Monday, and if ¬H, that was on Tuesday. (She'll replace *Tuesday* with *Wednesday* in the three-day variant.) Call this conditional *C*. Now think back to that last waking. At that time she is wondering various thoughts about the waking she is currently undergoing, a waking she will describe as "this waking". When she wakes on Wednesday, and thinks about "that waking", it is plausible she is thinking the very same kind of Fregean thought. She is thinking about the same thing, and thinking about it in the same kind of way. If that's right, then the big difference between her credences after the puzzle ends and her credences on the last waking are just that she comes to learn *C*. So on the last waking, her credence in *H* conditional on *C* should be ½, since when she learns *C* and nothing else, her credence in *H* becomes ½. And that entails (3), and similar reasoning generalises to all Sleeping Beauty problems to entail **Last Day**.[4]

The argument in Weatherson (2011) for (1) , and for **First Day**, involved a rather baroque variant of the example involving time travelers. And, as we've noted already, it also involves some fallacious equivocation. Before we get to that, it is worth noting two other arguments for the same conclusion, neither of which generalise to the Generalised Thirder Principle. Both arguments have contentious premises, but they are somewhat independent, so I hope that presenting both arguments will increase the number of people who agree with **FirstDay**.

The first argument is an argument from the Principal Principle. Consider the version of the Sleeping Beauty puzzle where the coin is tossed after Beauty goes back to sleep on Monday. (We didn't say so far when the coin is tossed, and it was consistent with everything we said that it is Monday night.) And assume that Beauty knows that the coin is tossed Monday. So when she wakes on Monday, she knows that the chance of *H* is ½ if it is Monday. That is, $Cr_1(Ch(H) = ½ \mid M) = 1$. The Principal Principle says that unless Beauty has inadmissible evidence, $Cr_1(H \mid Cr(H) = ½) = ½$. And plausibly she doesn't have any inadmissible evidence conditional on it being Monday. Putting these two together, we get $Cr_1(H \mid M) = ½$. And that kind of reasoning generalises to support **First Day**.

The argument here is similar to the original argument given for (1) in Elga (2000). Elga imagines a variant of the example where Beauty is told, sometime after she wakes up, that it is Monday, and argues that after that she should have credence ½ in *H*, and derives (1) from that. Halpern (2004) objects to this argument on the grounds that the possibility of Beauty being told what day it is undermines the indistinguishability of the wakings, and this undermines Elga's own argument for (2). I'm not sure Halpern is right, but in any case, this argument doesn't rely on any possibility of Beauty being told what day it is. I imagine that some people will object to the claim that Beauty has no inadmissible evidence. But it is hard to see what she knows which is inadmissible,

inadmissible. Or one might be tempted by a Reflection Principle based argument against any alternative credence. But such arguments seem to lead to odd results in general around Sleeping Beauty. It's best, I think, to stick with the clear intuition that on Wednesday her credence in *H* should be ½.

[4]Strictly speaking, all we've really shown is that Beauty's credences on the last day she wakes should satisfy **Last Day**. So if the focus of the puzzle is on her credences on the first day, all we've strictly speaking shown is that if *H*, then **Last Day** is true. I think it is plausible that **Last Day** should be independent of how the coin lands, but I admit that I don't have an argument against someone who wants to dispute this.

at least conditional on it being Monday. She *knows* that if it is Monday, the truth of H rests on a chance event that is yet to take place, and from which she is causally isolated. That looks to me like knowledge that she has no inadmissible evidence.

The second argument is a version of the the Technicolor Beauty argument in Titlebaum (2008). It relies on a variant of the example that drops the idea that the wakings are indistinguishable in a strong sense. I'll set up first what the idea behind the argument is, and then set out how it works. Assume that Beauty can think about M on Sunday. Follow Ross in using Cr_0 for Beauty's credences on Sunday. The same Principal Principle style argument we used above suggests that $Cr_0(H \mid M) = ½$. Indeed in this case the argument is even stronger, since everyone agrees that on Sunday, Beauty has no inadmissible evidence. But nothing happens to surprise Beauty between Sunday and Monday, so $Cr_1(H \mid M) = ½$ should be ½ as well, and we derive (1) from there.

There are a few problems with this argument. For one thing, the 'no surprise' premise goes by very quickly. More importantly, Beauty can't actually have M thoughts on Sunday. She can think to herself that Monday is Monday, or at least she could if she cared to think about the law of identity. But that's not the same thought as M. Remember, the guiding idea here is that contents are Fregean; it isn't easy to have the same thought as someone who thinks *This is Monday*. Something dramatic needs to happen to let Beauty have such a thought on Sunday, when she isn't in a position to make the same kind of demonstration as she is on Monday.

Here's one way the dramatic thing might happen. Change the example so that the wakings Beauty undergoes are not *phenomenally* indistinguishable. In fact, Beauty is told on Sunday that each waking will be in a brightly coloured room, and the colours will be different each day. As it happens, the room that will be used for Monday is red, though Beauty doesn't know that. Let RE be the proposition that one of Beauty's wakings will be in a red room, and RM be the proposition that she wakes Monday in a red room. Now clearly RE, on its own, is inadmissible evidence in the sense that $Cr_0(H \mid RE)$ need not be ½. After all, RE is probabilistic evidence that Beauty has more than one waking, since the more wakings she has, the more chance there is that one of them will be in a red room. On the other hand, RM does not look like inadmissible evidence. She has to wake up in some colour room or other on Monday; learning it is red doesn't change anything. So $Cr_0(H \mid RM)$ should be ½. And that's true even though RM obviously entails RE.

Now she wakes on Monday, and the room is red. What follows? Well, she now knows RE. And she can identify her current waking with the red waking she imagined (or at least could have imagined) on Sunday. So it is at least arguable that when on Monday she considers the thought *This waking is on Monday*, that's the very same thought she considers on Sunday by saying to herself *The waking in a red room is on Monday*. Making that last claim more plausible would require offering a more detailed theory of mental content than I have the space (or ability) to do here. For now I'm just going to take as a premise that there's a workable theory of mental content that types contents more finely than does a purely referential theory, but on which it is nevertheless the case that Beauty's demonstrative thought on Monday has the same content as her

descriptive thought (about the waking in the red room) does on Sunday.

Here is one way of thinking about that claim about content that may make it more plausible. (This idea is derived from the arguments in Jeshion (2002).) Imagine that on Sunday Beauty names the waking in a red room. She calls it 'Bluey'. She knows that 'Bluey' might not refer. That is, she knows that Bluey, like Vulcan and Sherlock Holmes, might not exist. But she nevertheless entertains detailed thoughts about Bluey. She wonders if Bluey will be on Monday, whether she'll be happy when Bluey happens, and so on. Now she wakes up on Monday, and sees the red walls. She says to herself, "This is Bluey". From that point on, it seems that she'd express the same thought with *This waking is* φ and *Bluey is* φ, and it seems she's express the same thought with *Bluey is* φ and *The waking in a red room is* φ. By appeal to transitivity of identity, and substituting a particular value for φ, we get that she expresses the same thought by saying *This waking is on Monday* as by saying *The waking in a red room is on Monday*.

If that claim about content is right, then *all* that happens on Monday is that Beauty learns that *RE* is true. She doesn't acquire the ability to think new thoughts, or to make fresh divisions in possibility space, the way that she does in the standard version of the puzzle. In the standard version of the puzzle, the demonstrative thought she considers on Monday, the one she would express by saying *This is Monday*, is not equivalent to anything she can think on Sunday. So when she wakes, she not only acquires some evidence, she acquires a new cognitive capacity. That doesn't happen here, which makes the calculations easier.

In particular, it lets us appeal to the following key fact. If E_2 entails E_1, and a particular update only involves conditionalising one's prior credences, then learning E_1 doesn't change the conditional credence of anything given E_2. That's a consequence of the following theorem. Let Pr be any probability function, and let $\Pr^+$ be the result of conditionalising that function on E_1. Then $\Pr(H \mid E_2) = \Pr^+(H \mid E_2)$. So if Beauty only learns *RE*, that doesn't change the conditional credence of anything given *RM*. In particular, it doesn't change the conditional credence of H given *RM*. So $Cr_1(H \mid RM) = ½$. And since M and *RM* are trivially equivalent, since Beauty can see the room is red, it follows that $Cr_1(H \mid M) = ½$, as required.

I suspect the main objection to this argument will be that adding the room colours makes a substantial change to the problem. The fact that ⅓ is the correct answer in this technicolour version of Sleeping Beauty, says the objector, is no reason to think that it is also the correct answer in the version where the wakings are phenomenally indistinguishable. But I think the objector will have a hard time making the case that phenomenal indistinguishability is epistemically significant unless they want to defend what Williamson (2007) calls the 'phenomenal conception of evidence'. As has been pointed out in prior work on Sleeping Beauty, the different wakings are not evidentially equivalent; when she wakes up and sees that *this* waking is happening, that's a piece of evidence she gets in some but not all wakings. (This point is made in Weintraub (2004) and Stalnaker (2008).) It might well be argued that this demonstrative evidence is in a sense symmetric; although she gets evidence that is different in some sense on the different wakings, the force of that evidence is the same. But that's still true in the

technicolour version of the problem as well. So I think, contra the objector, that this is a good argument for (1) .

But neither of those were the argument offered in Weatherson (2011). That argument involved a rather baroque modification to the puzzle. A time traveller films Beauty waking on Monday and travels back to Sunday to show Beauty the film. After she sees the film, and can think about the waking it depicts, the traveller tells her that he took it on Monday. He then wipes Beauty's memories of this telling, but not of the showing of the film. The argument then proceeds as follows.

1. On Sunday, she can think about her Monday waking in the same way as she thinks about it on Monday when she wakes, thanks to the time traveller's film.
2. On Sunday, the rational credence in H is ½.
3. If premise 1 is true, then on Monday, after she wakes, the only difference between her epistemic state then and her epistemic state after being told that the film was of Monday is that she no longer knows M.
4. If the only difference between two epistemic states is that in the first, an agent knows M and in the second she does not, then the rational credence of H given M in the second state equals the rational credence of H in the first state.

And from that $Cr_1(H \mid M) = ½$ was claimed to follow. Now there wasn't much of an argument for the first premise offered, and it might well be thought objectionable. It certainly relies on a liberal conception of sameness of 'ways of thinking'. But let's set that aside, because there is a much bigger problem with the argument. It hopelessly equivocates on the phrase 'On Sunday'. Let's distinguish the following four times that are all on Sunday.

- t_1 is before the time traveller turns up.
- t_2 is immediately after the time traveller shows Beauty the film.
- t_3 is immediately after the time traveller tells Beauty that the film is of Monday.
- t_4 is after the time traveller wipes Beauty's memories of that telling, but not of the showing of the film.

Now let's consider the first two premises in Weatherson's argument. The first premise is clearly false if 'On Sunday' refers to t_1, but arguably true if it refers to t_2, t_3 or t_4. The second premise is clearly true if 'On Sunday' refers to t_1, but much less plausible if it refers to any later time. If it refers to t_2 or t_4, it is arguably equivalent to the ½ answer to the original Sleeping Beauty problem, which makes it pretty useless in an argument for the ⅓ answer! If it refers to t_3, it is much too close to what we're trying to prove in arguing for **FirstDay**, so it is still argumentatively useless.

Ross argues that the style of argument we've been considering would, if it showed that (1) , and indeed more generally supported **First Day**, would show much more. In fact, it would show that for any two chance hypotheses h_i and h_j, and any k such that h_j is consistent with at least k wakings, that the following is true. (I'll again use p_{kj} to mean that this is the k'th waking and h_j is true.)

$$Cr_1(p_{1i} \mid p_{1i} \vee p_{kj}) = Ch(h_i)/(Ch(h_i) + Ch(h_j))$$

And from that we can derive the Generalised Thirder Principle, and hence countable additivity violations. That wasn't what was intended; the argument was only designed to work for the special case where $k = 1$, i.e., **First Day**. Now I think Ross is right in the following sense; there's just as good an argument in Weatherson (2011) for the above equation as there is for **First Day**. But that argument is no good for the reasons described above. Let's see where the same equivocation comes into Ross's telling of the story. I've changed Ross's notation a fair bit into notation I find easier to work with. Hopefully I haven't lost anything in the process. I'm also going to focus on the special case of Three-Day Sleeping Beauty, where h_1 is the coin lands heads and $h_{2\$}$is the coin lands tails, and on the case where $k = 2$. So what we're really going to look at is whether there's an argument in Three-Day Sleeping Beauty for this equation.

$$Cr_3(p_{11} \mid p_{11} \vee p_{22}) = Ch(h_1)/(Ch(h_1) + Ch(h_2)) = ½$$

Now as noted above, I think that this equation need not hold in Three-Day Sleeping Beauty; in the model I gave for it earlier, $Cr_3(p_{11}) = 1/5$, and $Cr_3(p_{22}) = 2/5$, so $Cr_3(p_{11} \mid p_{11} \vee p_{22}) = ⅓$. But let's see how Ross derives the ½ answer.

Again there is a time-traveller who shows Beauty a film. But this isn't necessarily a film of the first waking; the time traveller films the first waking if h_1, and the second waking if h_2. After seeing the film, Beauty is told this. So on Sunday, says Ross, Beauty's credences should satisfy these constraints. (I'm following Ross in using Cr_0 for the Sunday credences.)

$$^*Cr_0(h_1) = Cr_0(h_1 \mid p_{11})$$
$$Cr_0(h_1) = Ch(h^*_1)$$

There are other premises used, but these will be the crucial ones. They're crucial because there's no good reason to think that there's *any* time Sunday when Beauty's credences should satisfy both these equations. Before she sees the film, she can't even think about propositions like p_{11}, since she can't have singular thoughts about the waking it depicts. After she sees the film, there is no reason to think that her credences should align with chances. Causal contact with a time traveller who brings information that may well be about a time after the chance event, evidence whose existence may depend on the outcome of the chance event, is pretty much paradigmatically inadmissible evidence for the purposes of the Principal Principle. So after the film, there is no reason to think $Cr_0(h_1) = Ch(h_1)$.

I should stress that Ross doesn't *endorse* the equivocating premises here; he merely attributes them to Weatherson (2011), and fairly so. But I think once we see the equivocation we can see there is no fear the kind of argument used in Weatherson (2011) will lead to the Generalised Thirder Principle. That argument is too flawed to lead to anything. But there are plenty of other arguments for **First Day**, such as the two arguments offered here. And both of those arguments rely on distinctive features of the

first day. Most notably, they rely on the fact that for all we say in the setup of the problem, the first waking is before the chance event. So there's no reason to think they will have the problematic consequences that Ross finds in the argument for **First Day** in Weatherson (2011).

References

Elga, Adam. 2000. "Self-Locating Belief and the Sleeping Beauty Problem." *Analysis* 60 (4): 143–47. doi: 10.1093/analys/60.2.143.

Halpern, Joseph. 2004. "Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems." In *Oxford Studies in Epistemology*, 1:111–42. Oxford: Oxford University Press.

Jeshion, Robin. 2002. "Acquiantanceless *de Re* Belief'." In *Meaning and Truth: Investigations in Philosophical Semantics*, edited by Joseph Keim Campbell, Michael O'Rourke, and David Shier, 53–74. New York: Seven Bridges Press.

Ross, Jacob. 2010. "Sleeping Beauty, Countable Additivity, and Rational Dilemmas." *Philosophical Review* 119 (4): 411–47. doi: 10.1215/00318108-2010-010.

———. 2012. "All Roads Lead to Violations of Countable Additivity." *Philosophical Studies* 161 (3): 381–90. doi: 10.1007/s11098-011-9744-z.

Stalnaker, Robert. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

Titlebaum, Michael. 2008. "The Relevance of Self-Locating Beliefs." *Philosophical Review* 117 (4): 555–605. doi: 10.1215/00318108-2008-016.

Weatherson, Brian. 2011. "Stalnaker on Sleeping Beauty." *Philosophical Studies* 155 (3): 445–56. doi: 10.1007/s11098-010-9613-1.

Weintraub, Ruth. 2004. "Sleeping Beauty: A Simple Solution." *Analysis* 64 (1): 8–10. doi: 10.1093/analys/64.1.8.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Blackwell.

Published in *Philosophical Studies*, 2013, pp. 503-512.