

Everettian Confirmation and Sleeping Beauty: Reply to Wilson

Darren Bradley

ABSTRACT

In Bradley ([2011b]), I offered an analysis of Sleeping Beauty and the Everettian interpretation of quantum mechanics (EQM). I argued that one can avoid a kind of easy confirmation of EQM by paying attention to observation selection effects, that halvers are right about Sleeping Beauty, and that thirder cannot avoid easy confirmation for the truth of EQM. Wilson ([forthcoming]) agrees with my analysis of observation selection effects in EQM, but goes on to, first, defend Elga's ([2000]) thirder argument on Sleeping Beauty and, second, argue that the analogy I draw between Sleeping Beauty and EQM fails. I will argue that neither point succeeds.

- 1 *Introduction*
 - 2 *Background*
 - 3 *Wilson's Argument for $\frac{1}{3}$ in Sleeping Beauty*
 - 4 *Reply: Explaining Away the Crazy*
 - 5 *Wilson's Argument for the Breakdown of the Analogy*
 - 6 *Reply: The Irrelevance of Chance*
 - 7 *Conclusion*
-

1 Introduction

In my ([2011b]), I offered an analysis of Sleeping Beauty and the Everettian interpretation of quantum mechanics (EQM). I argued that one can avoid a kind of easy confirmation of EQM by paying attention to observation selection effects, that halvers are right about Sleeping Beauty, and that thirder cannot avoid easy confirmation for the truth of EQM. Wilson ([forthcoming]) agrees with my analysis of observation selection effects in EQM, but goes on to, first, defend Elga's ([2000]) thirder argument on Sleeping Beauty and, second, argue that the analogy I draw between Sleeping Beauty and EQM fails. I will argue that neither point succeeds.

After setting up the background in Section 2, I discuss the first point in Sections 3 and 4, and the second point in Sections 5 and 6. I conclude in Section 7.

2 Background

Consider the following two cases:

Quantum Wombat: Wombat is about to perform a spin measurement with possible outcomes Up and Down. Quantum mechanics says that Up and Down each has a chance of fifty percent. According to EQM, the universe will divide, so Wombat will have two future successors, one of whom will observe Up, and one Down. According to stochastic theory (ST), there will be only one future successor, who will observe either Up or Down, each with fifty percent probability. Wombat is unsure whether EQM or ST is correct, and assigns each a credence of fifty percent.

After branching and observing either Up or Down, what should Wombat's credence in ST be? Some say it should stay at $\frac{1}{2}$; call this 'no-easy-confirmation'. Others say it falls to $\frac{1}{3}$; call this 'easy-confirmation'.

Technicolour Sleeping Beauty¹: Beauty will be put to sleep on Sunday night and woken on Monday. A fair coin is tossed on Monday night. If the coin comes up Heads, Beauty will not be woken on Tuesday. If the coin comes up Tails, Beauty will be woken on Tuesday. Beauty's memory of her Monday experience will be erased on Monday night; so each waking is initially subjectively indistinguishable from every other. However, shortly after each waking Beauty will be shown either a Red or a Blue piece of paper. If Tails comes up, she will be shown Red on one day and Blue on the other, with a further fair coin determining on which day she'll be shown which colour²; if Heads comes up, she will be shown either Red or Blue on Monday, depending on the toss of a further fair coin. Beauty knows all this.

On waking and seeing a red piece of paper, what should Beauty's credence in Heads be? $\frac{1}{2}$ (so say halfers) or $\frac{1}{3}$ (so say thirders)?

In both cases, there are two possible worlds: a world with one successor (Heads/ST) and a world with two successors (Tails/EQM). And in both

¹ Wilson and I agree that modifying the original Sleeping Beauty problem by changing the coin toss to Monday doesn't change anything, so I'll discuss this variant. Other details we agree on will also be suppressed.

² There is a slip in Wilson's description of the case. He says, 'If Tails comes up, she will be shown Red on Monday and Blue on Tuesday'. He confirms (personal correspondence) this is not what he meant.

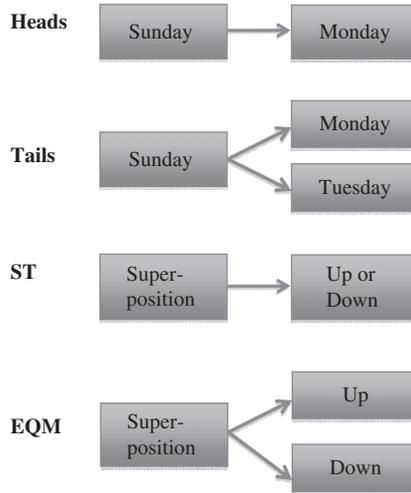


Figure 1. The analogy between Sleeping Beauty and EQM.

cases, the three possible observers have the same subjective experiences. The parallel can best be seen in [Figure 1](#).

I'll now briefly recap my ([2011b](#)) arguments, and highlight the points where Wilson disagrees.

In my ([2011b](#)), I made three claims that are relevant here. First, I gave an analysis of Quantum Wombat,³ which supports no-easy-confirmation. One aim of my argument was to deny the claim, which I attributed to Papineau and Durà-Vilà ([2009a](#), [2009b](#)), that the total evidence Quantum Wombat has on waking and observing Up is, 'there is a branch in which Up occurs'. Such evidence would indeed confirm EQM. Instead, I argued that the total evidence is something like, 'I learn that Up occurs in this branch by a random procedure' (where a random procedure would select each waking day with equal chance). Such evidence does not favour EQM over ST, so there is no easy confirmation. I'm gratified that Wilson agrees with this analysis.⁴

Second, I gave an analogous analysis of Sleeping Beauty that supports halvers. One aim of my argument was to deny the claim, which I attributed to Titelbaum ([2008](#)), that the total evidence Sleeping Beauty has on waking and seeing a red piece of paper is, 'there is a waking on which the red paper is observed'. Such evidence would indeed confirm Tails. Instead, I argued that

³ Though not using that name, which is Wilson's.

⁴ Though he does poke at a squishy bit in the argument regarding the principle of indifference and suggests a way to firm it up. I suggest a different way in my ([2012](#), p.160).

the total evidence is something like, ‘I learn that today is a red paper day by a random procedure’. Such evidence does not favour Tails over Heads, so there is no support for thirders.

Wilson agrees with this analysis too. However, he thinks there is a different route to the thirder position, that of Elga ([2000]), which is based on the principal principle (to be explained shortly). I will argue in Section 3 that this route doesn’t succeed either. So I will defend:

- (A) Elga’s ([2000]) argument for the thirder position, based on the principal principle, fails.

Third, I suggested in my ([2011b]) that due to the parallels between Quantum Wombat and Sleeping Beauty, the following conditional holds:

- (B) If you are a thirder in Sleeping Beauty, you are committed to easy confirmation of EQM.

Wilson denies (B).

I’ll explain Wilson’s denial of (A) in Section 3 and criticize it in Section 4. I’ll explain Wilson’s denial of (B) in Section 5 and criticize it in Section 6.

3 Wilson’s Argument for $\frac{1}{3}$ in Sleeping Beauty

Wilson agrees with me that the evidence Beauty has on seeing a coloured piece of paper doesn’t confirm Tails as some thirders claim. Instead, Wilson offers a different route to the thirder conclusion, via Elga’s ([2000]) argument.

To understand Elga’s argument, we need the principal principle and the concept of inadmissible evidence. The principal principle connects chance with rational belief. We can use Wilson’s locution (based on Lewis [1980]):

Principal principle (PP): Where an agent knows the chances and has no inadmissible evidence, the agent’s credences should match the chances.

And we can understand inadmissible evidence as:

Inadmissible evidence: Evidence that justifies an agent in having a credence that deviates from the known chances.

Wilson says his argument uses the following assumption:

The [...] assumption is that chance is the norm of credence: that in a situation where an agent knows the chances and has no inadmissible information, the agent’s credences should match the chances [PP]. Where a fair coin toss is in the future, an agent cannot have inadmissible information about it without the help of precognition or some other form of backwards causation. (Wilson [forthcoming])

In fact, there are two assumptions here. I agree with the first sentence—PP—and disagree with the second, which is a claim about what information

is inadmissible. Wilson's argument ([forthcoming]) needs both assumptions; it runs as follows:

Let Cr be Sleeping Beauty's credence function after waking but before being told what day it is:

- (i) $Cr(\text{Heads} \mid \text{Today is Monday}) = \frac{1}{2}$, by the PP,
- (ii) $Cr(\text{Heads} \mid \text{Today is Tuesday}) = 0$,
- (iii) $Cr(\text{Today is Tuesday}) > 0$,

Therefore, $Cr(\text{Heads}) < \frac{1}{2}$.

4 Reply: Explaining Away the Crazy

I deny (i). Wilson doesn't just apply PP; he assumes that 'today is Monday' is admissible evidence. It would be admissible evidence if 'precognition of some other form of backwards causation' were required for inadmissibility, as Wilson suggests ([forthcoming]). I claim that Beauty has inadmissible evidence without backwards causation. Though this doesn't come up in my ([2011b]), I gave a detailed argument in my ([2011a]), which is based on (Lewis [2001]) and which Wilson doesn't discuss. But he does say that the position defended there is 'crazy' and 'implausible'. So I will try to make the position more plausible.

I hold that Beauty should have credence $\frac{2}{3}$ that a future coin toss will land Heads:

- (i*) $Cr(\text{Heads} \mid \text{Today is Monday}) = \frac{2}{3}$, by the PP, correctly understood.

I agree that this is odd. But oddity is no objection by itself—Sleeping Beauty is in an odd position.⁵ What I hope to show is that the oddity I endorse is independently motivated. What follows in this section is a sketch of the argument of my ([2011a]), with an emphasis on making the position intuitively acceptable.

Consider the following case:

Imperfect crystal ball: Suppose you have an imperfect crystal ball. You can ask it about the result of any particular coin toss. Whenever the coin will land Heads, it shows you a picture of the coin showing Heads. But whenever the coin will land Tails, it shows you nothing. Suppose you ask the crystal ball about some particular toss, look into the crystal ball and it shows you nothing.

⁵ Titelbaum ([2012]) shows that even the double-halfer position (where Beauty's credence on waking and then learning that it's Monday should stay at one half) cannot avoid embarrassment.

Does this confirm Tails? Yes. One way to put this is to say that you have discovered an absence of evidence for Heads, and this confirms Tails. Another way to put this is to say that the evidence space was {image of heads, image of nothing}, and when one piece of evidence confirms a hypothesis (image of heads), the negation of that evidence (image of nothing) disconfirms the hypothesis.

Now imagine someone who looks at the image-free crystal ball and says: ‘The crystal isn’t working. So you have no inadmissible evidence and should follow the PP and believe Heads with fifty percent certainty’. They would be wrong. If they had correctly taken into account not just the absence of the image in the crystal ball but the whole evidence space, they would see that you have evidence for Tails.

I claim the same thing happens to Sleeping Beauty. Imagine her, after waking, about to be told what day it is.⁶ The evidence space is {today is Monday, today is Tuesday}. If she learns that today is Tuesday, this is conclusive evidence that the coin landed Tails. So if she fails to learn that today is Tuesday—if she learns that it is Monday instead—this is evidence that the coin lands Heads. So learning that it’s Monday is relevant to the future coin toss.

It is tempting to say, as the thirder does, ‘you have no inadmissible evidence and should follow the PP and believe Heads with fifty percent certainty’—that is, (i). But, I claim, that would be a mistake. If we take into account the whole evidence space, we can see that Beauty has evidence for Heads—that is, (i*). Therefore, we can conclude that (A): Elga’s ([2000]) argument for the thirder position fails.

5 Wilson’s Argument for the Breakdown of the Analogy

Let’s put that aside and grant that Wilson’s argument for the thirder position in Sleeping Beauty succeeds. I suggested in ([2011b]) that:

(B) If you are a thirder in Sleeping Beauty, you are committed to easy confirmation of EQM.

The argument for (B), in as much as I gave one, was simply that it seemed to follow from the parallels described at the beginning of this article. As I put it:

Thirders think that Beauty gains on waking some reason to believe that the possible world in which there are more observations (Tails) is more probable than the world with fewer (Heads). The analogous position

⁶ I assume here that on learning what day it is, Beauty should update by conditionalization. There is no memory loss during this period so no reason to deny conditionalization. Either way, Wilson gives no indication that he does.

regarding EQM is that we gain some reason after branching to believe that the possible world in which there are more observations (EQM) is more probable than the one with fewer (ST). As branching is happening all the time, it would follow that we have overwhelming evidence in favour of EQM! [...] If Thirders are to reject this easy evidence for [EQM], they owe us an explanation of where the disanalogy lies between Sleeping Beauty and EQM. ([2011b], p. 336; notation altered)

Wilson attempts to give an explanation of where the disanalogy lies. He argues that confirmation in EQM does not depend on a chance process, whereas confirmation in Sleeping Beauty does. As we saw above, Wilson's thirder argument relies on PP, which relies on a chance process generating Heads or Tails. As there is no chance process determining ST or EQM, the PP is inapplicable, so the thirder argument in Sleeping Beauty does not transfer to easy confirmation for EQM. Here's how Wilson puts it:

The disanalogy between Sleeping Beauty (SB) and Quantum Wombat (QW) is, on reflection, a straightforward one. Whether EQM or ST is true does not depend on any chance process, and Wombat knows that. In contrast, whether the coin lands Heads or Tails does depend on a chance process, and Beauty knows that. (Wilson [forthcoming])

6 Reply: The Irrelevance of Chance

The first thing to say is that coin flipping is really a non-chancy process—how a coin lands is determined by how exactly it was flipped and caught. However, let's grant Wilson that the coin flip in Sleeping Beauty is chancy.⁷

My main response is that it's implausible that the presence or absence of chance could make the difference Wilson needs it to. Let's distinguish the question of Beauty's reason for her prior probability from the question of whether she gains evidence that shifts the credence from the prior probability.⁸ That is, we should distinguish what generated $Pr(H)$ from whether $Pr(H|E) > Pr(H)$. In Sleeping Beauty, the PP just fixes the prior probability of H (the former question). However, the issue between halfers and thirders is whether Beauty receives new evidence on waking that shifts her credence (the latter question). Halfers say no; thirders say yes. This issue, I claim, isn't affected by what the prior is based on.

⁷ I'm grateful to a referee for stressing this point. The fact that we can grant this so easily indicates that chanciness doesn't play an important role. Meacham ([2008], p. 259) seems to express the consensus: 'Note that the Principal Principle only plays a superficial role in the argument for Elga's proposal. The Principal Principle sets our credences in heads and tails on Sunday to $(1/2)/(1/2)$. But the argument goes through equally well given any reason for $(1/2)/(1/2)$ credences in heads and tails on Sunday'.

⁸ It isn't prior relative to all evidence. It's prior relative to learning how the coin landed.

We can press the point by imagining a non-chancy variant of Sleeping Beauty. Wilson provides us with such a variant, in which uncertainty about the result of a fair coin toss is replaced by uncertainty about the truth of a mathematical proposition. Wilson has to say that the thirder arguments cannot be applied to the following case⁹:

Mathematical Sleeping Beauty (MSB): On Sunday night Beauty has credence $\frac{1}{2}$ that Fermat's last theorem is true. She will be awakened on Monday if the theorem is true, and on both Monday and Tuesday (again with her memories from Monday erased) if the theorem is false. Beauty knows all this. The puzzle is to say what credence Beauty should have on Monday in the proposition that Fermat's last theorem is true (call this proposition True.)

I claim that making the hypothesis non-chancy doesn't change anything—if you're a thirder for chancy Sleeping Beauty cases, you should be a thirder for Non-chancy Sleeping Beauty cases.

However, before we get to the details, we can improve the example, as this variant has some unwanted complications. First, MSB cannot be an ideal Bayesian agent as ideal Bayesian agents know all mathematical truths. Second, in Sleeping Beauty there is a long-run frequency argument for $\frac{1}{3}$, based on the fact that, in the long run, there will be twice as many Tails awakenings as Heads awakenings. This argument cannot be used in MSB, as the truth-value of a mathematical proposition can't vary between wakings. Wilson notes both worries and tries to use the second to drive a wedge between Sleeping Beauty (for which he endorses the $\frac{1}{3}$ answer) and MSB (for which he doesn't).

However, we can avoid both complications by letting the coin flip depend on a non-chancy contingent proposition. Here's a suggestion¹⁰:

Non-chancy Sleeping Beauty: On Sunday night, Beauty has credence $\frac{1}{2}$ that an even number of stars will be visible in total on Monday night. She will be awakened on Monday if there is an odd number, and on both Monday and Tuesday (again with her memories from Monday erased) if there is an even number. Beauty knows all this. The puzzle is to say what credence Beauty should have on Monday in the proposition that the number of stars is Even.

I claim that this still doesn't change anything: if you are a thirder about the original case you should be a thirder about this one. Wilson is committed to

⁹ Wilson pulls back from explicitly endorsing the halfer position for Mathematical Sleeping Beauty. He just argues that 'there are important disanalogies between MSB and SB' and 'this provides reason to doubt whether our two cases [MSB and SB] have a uniform solution [$\frac{1}{3}$]' (Wilson [forthcoming]).

¹⁰ If you think this is chancy, feel free to substitute another proposition you think is non-chancy.

being a thirder in the original Sleeping Beauty and a halfer in Non-chancy Sleeping Beauty. And, I claim, this difference is inexplicable. The only difference between the cases is what generated the priors, so we should not end up with a difference regarding confirmation.

Wilson does offer an argument that attempts to explain the difference between chancy and Non-chancy Sleeping Beauty. He points out that the principal principle argument for $\frac{1}{3}$, which I described in Section 3, cannot be applied to Non-chancy Sleeping Beauty cases. And of course he's right that it can't be.

However, this response is unsatisfactory for a couple of reasons. First, most thirder arguments do still apply to Non-chancy Sleeping Beauty. Wilson would have to hold that all of these fail, but Elga's succeeds.¹¹ Second, Non-chancy Sleeping Beauty seems to show that, intuitively, chancy cases in which the principal principle can be applied should get the same verdict as non-chancy cases in which the principal principle cannot be applied. Wilson has to explain away this intuition. Merely pointing out that the principal principle cannot be applied to non-chancy cases doesn't do this.

And if we do apply the halfer answer in Non-chancy Sleeping Beauty, we get problematic consequences of the kind that Wilson is keen to avoid. The halfer position in Non-chancy Sleeping Beauty means her credence in Odd is still $\frac{1}{2}$ after waking. And learning it's Monday then increases credence in Odd to more than $\frac{1}{2}$.¹² So Wilson is committed to the position that being told it's Monday is relevant to the number of stars that will be seen on Monday night! This seems no better than the claim that being told it's Monday is relevant to the toss of a coin on Monday night.

We can also put pressure on the distinction Wilson is trying to draw (between chancy and non-chancy processes) from the other direction. For example, imagine that God chose between ST and EQM by flipping a chancy fair coin.

Indecisive God: God cannot decide between creating a world with ST or one with EQM. So he creates a chancy fair coin and flips it: ST if Heads; EQM if Tails.¹³

¹¹ There is at least a *prima facie* case that each of the following thirder arguments can be extended to Non-chancy Sleeping Beauty; see (Arntzenius [2003]; Dorr [2002]; Draper and Pust [2008], Hitchcock [2004]; Horgan [2004]; Seminar [2008]; Titelbaum [2008]; and Weintraub [2004]).

¹² See Lewis ([2001]) for details. Here, with Wilson, I ignore the 'double-halfer response' whereby credence would stay at $\frac{1}{2}$.

¹³ This is similar to Wilson's God's gambling game (GGG). The (insignificant) difference is that in GGG, God is so indecisive that he flips a new coin every time a measurement is made to determine whether there will be more than one branch. Wilson points out that we don't get easy confirmation of GGG. But that isn't the issue. Wilson concedes that given GGG we would get easy confirmation of EQM (or its analogue for each branch). Similarly, he should concede that given Indecisive God, we would get easy confirmation of EQM. So he is committed to an

The PP can be applied here, so Wilson seems committed to the analogue of the thirder position, and easy confirmation of EQM. However, it's inexplicable that whether God chose his design of the world by chance or had a clear intention all along could make such a difference to our epistemic position.

So the distinction between outcomes generated by a chance process and outcomes generated by a non-chance process cannot do the work Wilson wants it to. So I maintain that: (B) If you are a thirder in Sleeping Beauty, you are committed to easy confirmation of EQM.

7 Conclusion

To sum up, I've defended two theses. First, that Elga's ([2000]) argument for being a thirder is unpersuasive and, second, that the presence of chance in Sleeping Beauty is irrelevant to whether being woken confirms Heads or Tails. Let me briefly connect this to two broader issues.

In the broader debate between thirders and halfers, I think there is no knock-down argument either way, but the most reasonable position still seems to me to be the halfer position of Lewis ([2001]), defended in my ([2011a]).

There is also a broader debate about the extent of the analogy between Sleeping Beauty and EQM, and specifically about whether thirders are committed to easy confirmation of EQM. Wilson has pointed out a disanalogy between Sleeping Beauty and EQM, but I have argued that the disanalogy isn't relevant to whether the agent receives confirmatory evidence.¹⁴ So the challenge stands: if thirders are to reject this easy evidence for EQM, they owe us an explanation of where the disanalogy lies.¹⁵

Funding

Support for this project was provided by a PSC-CUNY Award, jointly funded by The Professional Staff Congress and The City University of New York.

*Philosophy Department
The City College of New York
160 Convent Ave, NY 10031, USA
bradleydarren@gmail.com*

inexplicable difference in confirmation depending on whether God picked ST or EQM by a chancy or non-chancy process.

¹⁴ Wilson himself discusses other disanalogies pointed out by Papineau and Durà-Vilà ([2009a], [2009b]) and Peterson ([2011]), and argues that these disanalogies are not relevant; what Wilson says here seems reasonable to me.

¹⁵ Thanks to Alastair Wilson and an anonymous referee for helpful comments on earlier drafts.

References

- Arntzenius, F. [2003]: 'Some Problems for Conditionalization and Reflection', *Journal of Philosophy*, **100**, pp. 356–70.
- Bradley, D. [2011a]: 'Self-location Is No Problem for Conditionalization', *Synthese*, **182**, pp. 393–411.
- Bradley, D. [2011b]: 'Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty', *British Journal for the Philosophy of Science*, **62**, pp. 323–42.
- Bradley, D. [2012]: 'Four Problems about Self-locating Belief', *Philosophical Review*, **121**, pp. 149–77.
- Dorr, C. [2002]: 'Sleeping Beauty: In Defence of Elga', *Analysis*, **62**, pp. 292–6.
- Draper, K. and Pust, J. [2008]: 'Diachronic Dutch Books and Sleeping Beauty', *Synthese*, **164**, pp. 282–7.
- Elga, A. [2000]: 'Self-locating Belief and the Sleeping Beauty Problem', *Analysis*, **60**, pp. 143–7.
- Hitchcock, C. [2004]: 'Beauty and the Bets', *Synthese*, **139**, pp. 405–20.
- Horgan, T. [2004]: 'Sleeping Beauty Awakened: New Odds at the Dawn of the New Day', *Analysis*, **64**, pp. 10–21.
- Lewis, D. [1980]: 'A Subjectivist's Guide to Objective Chance', in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Berkeley: University of California Press.
- Lewis, D. [2001]: 'Sleeping Beauty: Reply to Elga', *Analysis*, **61**, pp. 171–6.
- Meacham, C. J. G. [2008]: 'Sleeping Beauty and the Dynamics of *De Se* Beliefs', *Philosophical Studies*, **138**, pp. 245–69.
- Papineau, D. and Durà-Vilà, V. [2009a]: 'A Thirder and a Everettian: A Reply to Lewis's "Quantum Sleeping Beauty"', *Analysis*, **69**, pp. 78–85.
- Papineau, D. and Durà-Vilà, V. [2009b]: 'Reply to Lewis: Metaphysics versus Epistemology', *Analysis*, **69**, pp. 89–90.
- Peterson, D. [2011]: 'Beauty and the Books: A Response to Lewis's Quantum Sleeping Beauty Problem', *Synthese*, **181**, pp. 367–74.
- Seminar, O. [2008]: 'An Objectivist Argument for Thirdism', *Analysis*, **68**, pp. 149–55.
- Titelbaum, M. G. [2008]: 'The Relevance of Self-locating Beliefs', *Philosophical Review*, **117**, pp. 555–606.
- Titelbaum, M. G. [2012]: 'An Embarrassment for Double-Halfers', *Thought*, **1**, pp. 146–51.
- Weintraub, R. [2004]: 'Sleeping Beauty: A Simple Solution', *Analysis*, **64**, pp. 8–10.
- Wilson, A. [forthcoming]: 'Everettian Confirmation and Sleeping Beauty', *British Journal for the Philosophy of Science*.