

Quantum Sleeping Beauty

Peter J. Lewis
plewis@miami.edu

The Sleeping Beauty paradox in epistemology and the many-worlds interpretation of quantum mechanics both raise problems concerning subjective probability assignments.¹ Furthermore, there are striking parallels between the two cases; in both cases personal experience has a branching structure, and in both cases the agent loses herself among the branches. However, the treatment of probability is very different in the two cases, for no good reason that I can see. Suppose, then, that we adopt the same treatment of probability in each case. Then the dominant ‘thirder’ solution to the Sleeping Beauty paradox becomes incompatible with the tenability of the many-worlds interpretation.

Consider first the many-worlds interpretation of quantum mechanics,² and in particular what happens when an observer measures the x -spin of a spin-1/2 particle whose state is an eigenstate of z -spin. According to the many-worlds interpretation, the observer *branches* into two successor observers, one of whom sees the result ‘spin up’ and the other of whom sees the result ‘spin down’. Call the original (pre-measurement) branch b_0 , the branch in which the observer sees ‘spin up’ b_1 , and the branch in which she sees ‘spin down’ b_2 . Compare this to a simplified version of the Sleeping Beauty paradox.³ Sleeping Beauty is woken up on Sunday, and told the following: “We will wake you for an hour on Monday, and for an hour on Tuesday, and on Monday night we will administer a drug that will cause you to forget the Monday waking.” Her room contains no indication of what day it is. Call the time at which she is awake on Sunday t_0 , on Monday t_1 , and on Tuesday t_2 .

Note the parallels between the two cases. In the many-worlds case, the agent at b_0 is straightforwardly psychologically continuous with the agent at b_1 , and straightforwardly psychologically continuous with the agent at b_2 ; however, the agents at b_1 and b_2 are not psychologically continuous with each other. Similarly in the Sleeping Beauty case, the agent at t_0 is straightforwardly psychologically continuous with the agent at t_1 and with the agent at t_2 , but the agents at t_1 and t_2 are not straightforwardly psychologically continuous with each other due to the memory erasure. That is, in each case the personal experience of the agent exhibits a branching structure.

Furthermore, note that in each case the agent gets lost in this branching structure. This is clearer in the Sleeping Beauty case; when she wakes up, she no longer knows whether it is Monday or Tuesday. That is, at t_1 and t_2 she is uncertain of the truth-value of some self-locating beliefs, such as “Today is Monday” (Elga 2000). She could use a probability measure to quantify this uncertainty; presumably she should assign a probability of 1/2 to “Today is Monday” on the basis of some kind of indifference principle. In the case of the many-worlds interpretation, the parallel uncertainty can be produced by supposing that the observer is blindfolded; at b_1 and b_2 , she knows that the measurement has taken place, but she doesn’t know whether the result is ‘up’ or ‘down’ (Vaidman 2002a). Again, she is uncertain of the truth-value of some self-locating

¹ I will drop the term ‘subjective’ from now on; all probabilities should be understood as subjective probabilities, unless specified otherwise.

² That is, an interpretation in the tradition of Everett (1957). See in particular Wallace (2002).

³ Elga (2000). I have modified Elga’s example to make the parallels between the two cases clear; I return to the original version below.

beliefs, such as “The result is ‘up’ in this branch” (Ismael 2003). And again, she could use a probability measure to quantify this uncertainty; she should assign a probability of 1/2 to the cited belief, on the basis of the Born rule.⁴

Hence in both the many-worlds case and in the Sleeping Beauty case, there is a branching structure to subjective experience, which induces a loss of self-location information, and the resulting uncertainty can be quantified using a probability measure. However, there is also a significant difference between the two cases, namely in the treatment of probability *before* the branching event. According to the many-worlds interpretation of quantum mechanics, the observer should assign a probability of 1/2 to each measurement outcome even at b_0 . That is, the treatment of probability in the many-worlds case is just as if exactly one of the two outcomes occurs, where each outcome has an objective chance of 1/2. But there is no analogous pre-branching probability assignment in the Sleeping Beauty case; there is no sense in which Sleeping Beauty at t_0 should assign a probability of 1/2 to each of Monday and Tuesday. To what beliefs could she assign such probabilities? “Today is Monday” has a probability of 0 at t_0 ; she knows that today is Sunday. “I will wake up on Monday” has a probability of 1 at t_0 ; she knows that she will wake up on Monday (or at least, that she is psychologically continuous with someone who will wake up on Monday). There are no obvious candidates for a belief about Monday to which Sleeping Beauty should assign a probability of 1/2 at t_0 . Hence the treatment of probability in the Sleeping Beauty case is *not* just as if she is woken up on exactly one of Monday and Tuesday with an objective chance of 1/2 each.

Of course, similar considerations can be raised in the many-worlds case, too. To what beliefs, at b_0 , can the observer assign a probability of 1/2? “The result is ‘up’ in this branch” has a probability of 0 at b_0 , since this branch (b_0) contains no measurement results. “I will see the ‘up’ result” has a probability of 1, since the observer knows that she will see this result (or at least, that she is psychologically continuous with someone who will). Again, there are no obvious candidates for a belief concerning the results to which she should assign a probability of 1/2 at b_0 .

Nevertheless, many authors have argued that a pre-measurement probability assignment of 1/2 to the two results is appropriate in the many-worlds case. Saunders (1998) and Wallace (2005) argue that the branching structure of the observer’s experience makes the observer genuinely uncertain concerning what will happen to her; hence it makes sense for her to assign a probability of 1/2 to “I will see the ‘up’ result” at b_0 . Vaidman (2002a) argues that even though the observer is only genuinely uncertain after the measurement, nevertheless she should act as if she is uncertain even before the measurement, and this underwrites an effective probability assignment of 1/2 to “I will see the ‘up’ result” at b_0 . Papineau (2004) and Greaves (2004) argue that a probability assignment need have nothing to do with uncertainty; a probability assignment of 1/2 to “I will see the ‘up’ result” at b_0 indicates how much the observer cares about her successor at b_1 relative to her successor at b_2 .

My goal here is neither to defend nor to attack these argument strategies, but merely to note that they all remain controversial, and that if any such strategy works, it ought to work just as well in the structurally similar Sleeping Beauty case. That is, given the parallels between the two cases, then all other things being equal, we should expect the two cases to be covered by a uniform account of probability. But as it stands, the treatments of pre-branching probability are

⁴ The Born rule equates the probability of a branch with the square of its quantum mechanical amplitude. It might strike readers as suspicious that the justification for the probability assignment differs between the two cases; I return to this issue below.

not parallel. This leaves us with two options—applying the treatment of pre-branching probability from the Sleeping Beauty case to the many-worlds case, and applying the many-worlds treatment to the Sleeping Beauty case.

According to the first option, the treatment of probability in the Sleeping Beauty case is correct; Sleeping Beauty should assign a probability of $1/2$ to both “Today is Monday” and “Today is Tuesday” at t_1 and at t_2 , but she should assign a probability of 1 to both “I will wake up on Monday” and “I will wake up on Tuesday” at t_0 . Applying this to the many-worlds case, the (blindfolded) observer should assign a probability of $1/2$ to both “The result is ‘up’ in this branch” and “The result is ‘down’ in this branch” at b_1 and b_2 , but she should assign a probability of 1 to both “I will see the ‘up’ result” and “I will see the ‘down’ result” at b_0 . The trouble with this option is that it is fatal to the many-worlds interpretation of quantum mechanics; the observer’s pre-measurement predictions contradict the Born rule, which forms the empirical heart of quantum mechanics. That is, quantum mechanics requires that non-trivial probabilities be assigned to measurement results before the measurement has occurred, and if the many-worlds interpretation cannot deliver such probabilities, then it is untenable as an interpretation of quantum mechanics. Indeed, this is precisely why the authors cited above take such pains to try to justify pre-measurement probability assignments in the many-worlds interpretation.

So let us consider the second option, namely that the treatment of probability by the advocates of the many-worlds interpretation is correct. According to this option, the quantum observer should assign a probability of $1/2$ to both “I will see the ‘up’ result” and “I will see the ‘down’ result” at b_0 , so analogously, Sleeping Beauty should assign a probability of $1/2$ to both “I will wake up on Monday” and “I will wake up on Tuesday” at t_0 . That is, in both cases the branching process is treated as if it were an instance of objective chance, where just one branch occurs. This is highly counter-intuitive in the Sleeping Beauty case; as Elga notes, on Sunday Sleeping Beauty was “already certain that [she] would be awakened on Monday” (2000, 145). But perhaps our intuitions here are wrong, and perhaps one of the arguments mentioned above for the many-worlds case can show why they are wrong.

However, more is at stake here than our intuitions. To see this, consider the full Sleeping Beauty paradox, rather than the simplified version considered so far. On Sunday, Sleeping Beauty is told that a coin will be tossed. If the coin comes up tails, then she will be woken up on Monday and on Tuesday with memory erasure in between, as in the simplified version. If the coin comes up heads, then she will be woken up on Monday alone. (It doesn’t matter for present purposes if the coin toss is interpreted as a classical event described by objective chance or as a quantum event described by many-worlds branching.) The question at the heart of the paradox is what probability Sleeping Beauty should assign to ‘heads’ when she wakes up at t_1 . The dominant (‘thirder’) view is that the answer is $1/3$ (Elga 2000); the minority (‘halfer’) view is that the answer is $1/2$ (Lewis 2001).⁵

Suppose we impose the treatment of probability required for the tenability of the many-worlds interpretation on the full Sleeping Beauty paradox. Recall that the treatment of probability here is as if each branch has an objective chance of $1/2$ of occurring. In the Sleeping Beauty case, this means we can treat the situation as if Sleeping Beauty is told that when the coin comes up tails, she will be woken up either on Monday or on Tuesday, with equal (objective) probabilities. That is, we can treat the situation as if there are two coin tosses, where Sleeping Beauty will be woken up on Monday (but not Tuesday) if the first coin comes up heads, on Monday (but not Tuesday) if the first comes up tails and the second comes up heads, and on

⁵ Evidence for the widespread acceptance of the thirder solution can be found in White (2006).

Tuesday (but not Monday) if both come up tails. But if this is the right way to analyze the situation, then the paradox dissolves, and the halfer solution is clearly correct. After all, the motivation for the thirder solution is Sleeping Beauty's rational expectation at t_0 that she will be woken up twice as often if the (initial) coin-toss comes up tails as if it comes up heads. But on the current understanding of Sleeping Beauty's rational expectations, there is no such asymmetry; whatever the result of the (initial) coin-toss, she expects to be woken up once, although if the result is tails then she is uncertain as to which day she will see.

This analysis is straightforwardly applicable if the account of probability in the many-worlds interpretation involves genuine pre-branching uncertainty (as Saunders (1998) and Wallace (1995) maintain). Given such uncertainty, Sleeping Beauty should treat equal-amplitude branching just like a classical coin-toss, and hence she should view her situation as just like the case of two consecutive coin-tosses just described. Put another way, if Sleeping Beauty is genuinely uncertain at t_0 about whether, given that the (initial) coin-toss comes up tails, she will be woken up on Monday or on Tuesday, then her epistemic situation doesn't change between t_0 and t_1 . All that changes is her point of view; at t_1 , she is uncertain whether, given that the coin came up tails, she *is currently* awake on Monday or on Tuesday. But without a change in epistemic situation, there is no occasion for her to change the probability assigned to 'heads' between t_0 and t_1 , and at t_0 that probability is clearly 1/2.

But even if the account of probability in the many-worlds interpretation does not involve pre-branching uncertainty, all authors agree that the account must underwrite behavior exactly as if there was genuine uncertainty, in order for it to count as an account of *probability* (Greaves 2004, 442; Papineau 1996, 238; Vaidman 2002a). But in that case, Sleeping Beauty must behave exactly as if she is uncertain whether, given that the coin comes up tails, she will be woken up on Monday or on Tuesday. And part of behaving that way, as has just been argued, is being willing to bet on 'heads' at even odds at t_1 . Hence if the treatment of probability by the advocates of the many-worlds interpretation is correct, in any of its forms, then the dominant thirder solution to the Sleeping Beauty paradox must be rejected in favor of the halfer solution.⁶

The argument here assumes that there is no disanalogy between the many-worlds case and the Sleeping Beauty case that could justify different treatment of probabilities. But this might be challenged. For example, in the many-worlds case, the probabilities are related to the amplitude of the branch via the Born rule, and there is nothing analogous to branch amplitude in the Sleeping Beauty case. But note that this disanalogy is unrelated to the question of whether the relevant probabilities can be assigned prior to the branching event, which is the question at issue here. If there are further disanalogies between the two cases, the argument needs to be made. Otherwise, the parallels between the cases constrain your options: If you are a thirder, you must reject the many-worlds interpretation, and conversely, if you accept the many-worlds interpretation you must be a halfer.

References

- Elga, Adam (2000), "Self-locating belief and the Sleeping Beauty problem", *Analysis* 60: 143–147.
- Everett, Hugh III (1957), "'Relative state' formulation of quantum mechanics", *Reviews of Modern Physics* 29: 454–462.

⁶ Note that this is at odds with the conclusion of Vaidman (2002b), who maintains that the account of probability required for the many-worlds interpretation entails the thirder solution.

- Greaves, Hilary (2004), "Understanding Deutsch's probability in a deterministic multiverse", *Studies in History and Philosophy of Modern Physics* 35: 423–456.
- Ismael, Jenann (2003), "How to combine chance and determinism: Thinking about the future in an Everett universe", *Philosophy of Science* 70: 776–790.
- Lewis, David (2001), "Sleeping Beauty: Reply to Elga", *Analysis* 61: 171–176.
- Papineau, David (1996), "Many minds are no worse than one", *British Journal for the Philosophy of Science* 47: 233–241.
- (2004), "David Lewis and Schrödinger's cat", *Australasian Journal of Philosophy* 82: 153–169.
- Saunders, Simon (1998), "Time, quantum mechanics and probability", *Synthese* 114: 373–404.
- Vaidman, Lev (2002a), "Many-worlds interpretation of quantum mechanics", in Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
<http://plato.stanford.edu/archives/sum2002/entries/qm-manyworlds/>.
- (2002b), "Probability and the many worlds interpretation of quantum mechanics" in A. Khrennikov (ed.), *Quantum Theory: Reconsideration of Foundations*. Växjö, Sweden: Växjö University Press, 407–422.
- Wallace, David (2002a), "Worlds in the Everett interpretation", *Studies in History and Philosophy of Modern Physics* 33: 637–661.
- (2005), "Epistemology quantized: Circumstances in which we should come to believe in the Everett interpretation", <http://philsci-archive.pitt.edu/archive/00002368/>. Forthcoming in *British Journal for the Philosophy of Science*.
- White, Roger (2006), "The generalized Sleeping Beauty problem: A challenge for thirders", *Analysis*, forthcoming.