

# Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty

D. J. Bradley

---

5

## ABSTRACT

Sometimes we learn what the world is like, and sometimes we learn where in the world we are. Are there any interesting differences between the two kinds of cases? The main aim of this article is to argue that learning where we are in the world brings into view the same kind of observation selection effects that operate when sampling from a population. I will first explain what observation selection effects are (Section 1) and how they are relevant to learning where we are in the world (Section 2). I will show how measurements in the Many Worlds Interpretation of quantum mechanics can be understood as learning where you are in the world via some observation selection effect (Section 3). I will apply a similar argument to the Sleeping Beauty Problem (Section 4) and explain what I take the significance of the analogy to be (Section 5). Finally, I will defend the Restricted Principle of Indifference on which some of my arguments depend (Section 6).

10

15

### 1 *Selection Effects*

#### 1.1 *Biased procedure*

20

#### 1.2 *Random procedure*

### 2 *Centred Propositions and Selection Effects*

### 3 *Many Worlds*

### 4 *Sleeping Beauty*

### 5 *Significance of the Analogy*

25

### 6 *The Restricted Principle of Indifference*

---

## 1 Selection Effects

Whenever a sample is drawn from a population, some particular method must be used. This method is the selection procedure. The effect this has on the inference is the (observation<sup>1</sup>) selection effect. Eddington's ([1939])<sup>2</sup> classic example involves fishing with a net. If we catch a sample of fish from a lake, and all the fish in the sample are bigger than six inches, this appears to confirm the hypothesis that all the fish in the lake are bigger than six inches. But if we then find out that the net used cannot catch anything smaller than six inches due to the size of its holes, the hypothesis is no longer confirmed. So the inference depends on the method of obtaining the sample, i.e. on the selection procedure. There are countless types of selection procedure, but there are only two that we will need.

A *random* procedure is one where each member of the population has an equal chance of being selected for the sample.

A *biased*<sup>3</sup> procedure is one where, if any member of the population has a certain property  $p$  then it is selected for the sample. We will say that the procedure is biased towards  $p$  in such a case. I will be repeatedly referring back to this definition, so it is worth putting more succinctly. If a procedure is biased towards  $p$  then: if  $p$  is instantiated then  $p$  is observed.

To get a grip on these selection procedures, let's consider a simple thought-experiment that has the same structure as the cases we will look at.<sup>4</sup> Suppose you are faced with an urn containing either one ball or two, depending on the result of a fair coin toss (two if Tails, one if Heads). If Tails lands, one big ball and one small ball will be placed in the urn. If Heads lands, another fair coin will be flipped to determine if the single ball will be big or small. (For future reference, Heads and Tails correspond to uncentred possible worlds. Balls 1 and 2 correspond to centres within those worlds.)

<sup>1</sup> I will henceforth leave 'observation' implicit.

<sup>2</sup> Selection effects play an important role in Horwich's ([1982]) analysis of the ravens paradox and Jackson's analysis of the grue problem (Jackson [1975]; Jackson and Pargetter [1980]; Godfrey-Smith [2003]). Its most famous application is the Monty Hall problem (vos Savant [1997]). See (Hutchison [1999]; Bostrom [2002]) for systematic discussions. I use selection effects in (Bradley [2009], [2010]). See also (Freund [1965]; Shafer [1985]; Bovens and Ferreira [2010]; and Kotzen [unpublished]).

<sup>3</sup> These types are not exhaustive. I have used a precisification of the ordinary language term 'bias' that is simple and fits Sleeping Beauty and Many Worlds. It could be more precisely expressed as 'maximal bias'. Eddington's example fits a slightly different kind of bias: *only* objects with  $p$  are selected for the sample. I used this definition in (Bradley [2009]).

<sup>4</sup> Groisman ([2008]) also uses an urn analogy to model Sleeping Beauty, but for a very different purpose.

	<b>Ball 1</b>	<b>Ball 2</b>
<b>Heads</b>	Big or small	
<b>Tails</b>	Big	Small

5 A sample of one is taken from the urn. It turns out to be small.

10 How should we describe the evidence? One thing that has been learnt is ‘there was at least one small ball in the urn’. To express this in a similar way to the problem cases that follow, we can say ‘there is a ball which is small’.

$E$  = There is a ball which is small.

15 But this may not be the total evidence. We may also know the procedure by which we came to learn that there is a small ball. We’ll assume throughout that the agent has full and certain knowledge of the procedure. Let’s run through the effects of a biased and random procedure. (We’ll see that whereas in Eddington’s fishing case the hypothesis in question would be confirmed if the procedure were random, in this urn case the hypothesis in question would be confirmed if the procedure were biased. This is not hugely significant, but I mention it to avoid confusion.)

## 20 1.1 Biased procedure

First assume the procedure is biased towards smallness.

$E_B$  = I learnt that there is a small ball by a procedure biased towards smallness.

25 We could make this vivid by imagining that a small hole is opened and the urn shaken until a ball comes out. Big balls won’t fit, so a small ball will be observed whenever a small ball exists.<sup>5</sup> Recall that the probability that a small ball exists given Heads is  $1/2$  and the probability that a small ball exists given Tails is 1. So the probability that a small ball is selected given Heads is  $1/2$  and the probability that a small ball is selected given Tails is 1;  $P(E_B|H) = 1/2$  and  $P(E_B|T) = 1$ . Assuming the standard definition of confirmation— $E$  confirms  $H$  iff  $P(E|H) > P(E|\neg H)$ —(Hacking [1965]) it follows that  $E_B$  confirms Tails.

35 It will be important in what follows that Tails isn’t automatically confirmed whenever the procedure is biased. Tails is confirmed when a procedure that is biased towards  $p$  selects an object with  $p$ . If instead a procedure that is biased towards  $p$  doesn’t select  $p$ , then Tails is disconfirmed. For example, if a procedure that is biased towards small balls fails to select a small ball, then there

<sup>5</sup> This procedure is such that an object is selected for the sample if and only if it is small. This is a stronger bias than occurs in Sleeping Beauty and Many Worlds, as it is incompatible with observing a large ball.

must be no small balls in the urn. There must be just one ball in the urn (which is large), so Heads must have landed.

## 1.2 Random procedure

Now suppose a large hole is opened and the urn thoroughly shaken until a ball comes out, so that every ball has an equal chance of being selected.

$E_R = I$  learnt that there is a small ball by a random procedure

If Tails, there is one small ball and one large ball in the urn, so  $P(E_R | T) = 1/2$ . If Heads, there is a 50% chance of the urn containing a large ball and a 50% of the urn containing a small ball, so  $P(E_R | H) = 1/2$ .  $P(E_R | T) = P(E_R | H)$  so neither Heads nor Tails is confirmed.

When the procedure was biased towards small balls,  $P(E_B | T)$  was 1, as a small ball was certain to be in the urn (due to Tails), and that small ball was certain to be selected (due to the bias). But now that a ball is selected at random, a small ball is no more likely to be observed than a big ball. So  $P(E_R | T) = 1/2$ .

In both cases, the agent has learnt that there was a small ball in the urn. But the procedure by which this has come to be learnt affects the inferences we can draw. We can take the procedure into account by conditionalizing on ‘I learnt  $E$  by procedure  $p$ ’, as opposed to just ‘ $E$ ’. The distinction is between a proposition being true and being learnt. Bayesians generally assume that  $E$  is learnt whenever it is true.<sup>6</sup> This inference from ‘ $E$  is true’ to ‘ $E$  is learnt’ is equivalent to assuming the procedure is biased towards  $E$ , whatever  $E$  turns out to be.

We will see that in the Many Worlds Interpretation and Sleeping Beauty case, it is natural to assume the procedure is biased in this way. But once the procedure is brought to attention, it is clear that we should not make this assumption. Explaining this point is the main aim of this article. It follows that a problem of confirmation in Many Worlds is solved (Greaves [2004], [2007]) and an argument for the Thirder position in Sleeping Beauty (Titelbaum [2008]; Papineau and Durà-Vilà [2009b]) is shown to be problematic. Before moving on to this, I want to briefly explain the connection between selection effects and centred propositions.

## 2 Centred Propositions and Selection Effects

(This section can be skipped without loss of continuity.) Frege’s attempt to model our beliefs using uncentred propositions famously floundered on

<sup>6</sup> To take two prominent examples, Hacking ([1967], p. 316) treats the inference from the truth of  $p$  to learning  $p$  as a ‘trifling idealization’ and Howson and Urbach ([2006], p. 54) make use of an ‘omniscient oracle’.

indexicals. A man lost in a library isn't helped by knowing merely the uncentred proposition that 'agent A is at the south end of corridor F, level 3, at time t' because agent A could be someone else. The man needs to know that *he* is at that position *now* before he knows where he should go (Perry [1979]). Lewis ([1979]) gives the example of two Gods who know every uncentred proposition concerning their world, but each is still uncertain which of the two Gods he is. There is no way to describe what they are uncertain about using only uncentred propositions. We also need centred propositions. A centred proposition tells the agent when or where she is located. Sentences containing indexicals, such as 'I'm in London' or 'it is now midnight', express paradigm examples of centred propositions. Furthermore, sentences which don't contain indexicals, but which need to be relativized to a time or place in order to determine a truth-value, such as 'Obama is President', express centred propositions.

Lewis (*ibid.*) asked what happens when centred propositions are added to the Bayesian framework and answered 'not much'. But this answer raises two issues. The first is that beliefs represented by centred propositions change in ways that beliefs represented by uncentred propositions don't. This is because centred propositions can change in truth-value, so rational belief can change in virtue of tracking that truth-value. For example, the belief that today is Monday disappears as midnight strikes and is replaced by the belief that today is Tuesday. It is uncontroversial that the Bayesian framework—conditionalization in particular—must be modified to incorporate this kind of belief change. Various theories have been proposed (Meacham [2008]; Titelbaum [2008]; Schwarz [unpublished]) but I won't say anything about this kind of belief change here (my suggestion is in Bradley [forthcoming]).

The second issue arises in cases where the agent is initially *uncertain* about some self-locating fact, and then discovers it. This happens every time you look at a map to discover where you are, or look at a clock to learn the time.<sup>7</sup> Note that this is a different kind of case to that of the previous paragraph, for in this case the change in belief is not due to a change in the truth value of the object of belief. (The authors mentioned in the previous paragraph deal with these two issues together, but I think it best to separate them.) I will try to demonstrate that this second way of acquiring a self-locating belief does not present a problem for conditionalization. Instead, the complications in such cases are caused by selection effects. Selection effects are especially important to such cases because learning centred propositions has a degree of freedom which is analogous to the degree of freedom we have in selecting a sample.

<sup>7</sup> This is also the kind of learning that occurs when you learn who you are in the Doomsday argument (Leslie [1989]).

For example, suppose you are uncertain which of two regions of space you are in. It might be that you are equally likely to be in either, in which case each has a 50% probability of being observed by you. We can model this as a case in which there is a random selection procedure regarding which region of space you observe. But now suppose instead that one of the regions of space is life-permitting and the other is not. We can model this case as one in which the selection procedure is biased towards observing the life-permitting region of space. So we can see that when a centred proposition is learnt, we have to ask: is there a reason I am in one location rather than another? That is, what is the selection procedure by which I have observed one location rather than another?

I am not saying that selection effects only exist, or are only relevant, when a centred proposition is learnt. Quite the opposite—selection effects always exist, and are always relevant, for any piece of evidence, be it centred or uncentred.<sup>8</sup> The connection between selection effects and centred evidence that is most important for our purposes is that learning uncentred evidence is naturally modelled using a procedure whereby if  $p$  is instantiated then  $p$  is observed i.e. by a biased procedure. And this is not true of learning centred evidence.

For example, suppose that Bob is the only object in the population. Then any selection procedure that selects an object will select Bob. So any properties Bob has will be observed, and we can model this using a procedure that is biased towards whatever is observed. (If  $p$  is instantiated,  $p$  is observed.) We often learn uncentred propositions in an analogous way. If I am about to look at a die to see what number it landed on, then I will see the number it landed on. So we can model learning uncentred propositions using a procedure that is biased towards whatever property is observed (e.g. if a six is instantiated then a six is observed). However, if there have been numerous dice rolls, and I won't observe all of them, then there may be a six in the population that I don't observe. So the procedure cannot be modelled as biased towards whatever is observed. Similarly, when evidence is centred, we cannot automatically model the procedure as biased towards whatever is observed some of which I will observe and some of which I won't, and I am more likely to observe some numbers than others.

I will now argue that in both the Many Worlds case and Sleeping Beauty we must resist the temptation to treat the procedure as biased towards whatever is observed. I will argue that we should model both cases as learning a centred

<sup>8</sup> For example, you might learn that a particular raven, Bob, is black by being told by someone whose procedure is to tell you that Bob is black if and only if not all ravens are black. You have learnt uncentred evidence, yet the selection effect plays an important role in the inferences that can be drawn.

proposition via a random procedure and in both cases neither of the competing hypotheses is confirmed.<sup>9</sup>

### 3 Many Worlds

Suppose we are comparing two mutually exclusive and exhaustive interpretations of quantum mechanics. The stochastic theory (ST) says that the world is chancy—for a particle in a superposition of Up and Down that is about to be measured, ST says there is a certain chance of ‘up’, and a certain chance of ‘down’. Suppose that chance is 50%. The Many Worlds Interpretation (MWI) says that the world is branchy—the world will split into ‘up’ and ‘down’ branches when the measurement is made. Suppose the branches have a weight<sup>10</sup> of 50%. Note that ST and MWI are two possible worlds and the branches are centres within those worlds. In ST, there is only one branch; in MWI, there are two.

15

	<b>Branch 1</b>	<b>Branch 2</b>
ST	Up or Down	
MWI	Up	Down

(Note the similarity to the urn example.) Suppose an agent makes an observation of Up. Let’s first express that evidence as an uncentred proposition, without making reference to the selection effect, to see how this leads to trouble.

$E$  = There is a branch in which Up occurs.

(An obvious problem appears to be that  $E$ , by referring to a branch, presupposes that MWI is true. But if ST is true, there is only one branch, and  $E$  is referring to that one.)

Does  $E$  confirm MWI? This amounts to: is  $P(E|MWI) > P(E|ST)$ ? Yes. MWI says that both Up and Down will occur, so  $P(E|MWI) = 1$ , while  $P(E|ST)$  is  $1/2$ .

It looks like our evidence confirms MWI. And notice that it didn’t matter that Up was observed rather than Down. Both outcomes had a probability of 1 given MWI, both outcomes had a probability of  $1/2$  given ST, so both must confirm MWI. We get the result that any possible evidence confirms MWI.

This result is absurd. But it is not surprising; as MWI says that every outcome occurs, any outcome that isn’t entailed by ST will confirm MWI. So it

<sup>9</sup> This does not imply that evidence learnt by a random method can never confirm any hypothesis. Classical statistics is built on random sampling, e.g. Eddington’s fishing case above.

<sup>10</sup> ‘Weight’ is a rather mysterious entity that is required by MWI to fit the Born probabilities, and can be thought of as a measure of existence. Unequal weights bring in new complications which I will not go into in detail in this article. See footnote 14.

appears that every observation confirms MWI. This line of reasoning is at the root of what Greaves ([2007]) calls Naïve Conditionalization. She rejects (correctly) Naïve Conditionalization due to its absurd consequences, but she doesn't say what went wrong with the reasoning, and not everyone is convinced that it is wrong. Price ([unpublished]) suggests instead that either MWI is wrong, or that conditionalization can't be combined with MWI. The apparent incompatibility of MWI and conditionalization has been considered a fundamental problem with MWI. Until we have a diagnosis of what has gone wrong with Naïve Conditionalization, the suspicion will remain that attempts to give an alternative theory of confirmation in MWI are *ad hoc*. I will argue that all that has gone wrong is that incorrect conditional probabilities have been used. Naïve Conditionalization fails to use centred propositions and the observation selection effects that come with them.<sup>11</sup>

According to MWI, the evidence learnt when Up is observed is not the uncentred proposition that there is a branch in which Up occurs, but the centred proposition that Up occurs in *this* branch. The agent has learnt where he is—the Up branch. So the first thing we have to do to correct Naïve Conditionalization is express the evidence learnt as a proposition that is centred given MWI:

$E_c = \text{Up occurs in this branch}$

Notice that  $E_c$  only expresses its intended meaning if expressed *after* branching. The situation being modelled is that of an agent after branching but before learning which branch he is in.<sup>12</sup> The question at hand is to determine the likelihoods  $P(E_c | \text{MWI})$  and  $P(E_c | \text{ST})$  for such an agent.

The second thing we have to do to correct Naïve Conditionalization is put in the correct selection effect. The selection effect is trivial given ST — you simply observe whatever happens in the one 'branch' that exists. If the particle collapses to Up, Up will be observed. If the particle collapses to Down, Down will be observed. We are assuming each has a 50% chance, so  $P(E_c | \text{ST}) = 1/2$ . However, in order to determine the likelihood  $P(E_c | \text{MWI})$ , we have to ask: given that both outcomes occur, what should your subjective probability be that you—the in-branch agent you happen to be<sup>13</sup>—are in a branch in which Up

<sup>11</sup> Is my position incompatible with Greaves' quasi-conditionalization, or does it render it otiose? Neither. I am only concerned with the belief change between ( $t_1$ ) after the branching but when one is still uncertain which branch one is on, and ( $t_2$ ) after learning which branch one is on. Greaves is concerned with how we should update from ( $t_0$ ) before the branching has occurred. This is more difficult because before branching one cannot be uncertain which branch one is on, so there is no appropriate conditional probability to be used in conditionalization. (I assume the agent knows whether or not branching has occurred.)

<sup>12</sup> Wallace ([2006]) worries that the time between branching and the agent learning which branch they are in may be too short for them to be uncertain. But this appears to be just an instance of the problem of old evidence (Glymour [1980]).

<sup>13</sup> One might object that the only way to individuate 'you' is as 'the inhabitant of the Up branch', and argue that this has a chance of 1. But the probabilities are subjective, not objective. Even if

occurs? The answer to this question will depend on whether there is any bias in the selection procedure.

Suppose first that the procedure is biased towards Up branches. Recall this means that if there is an Up branch, then it is observed. We could make this  
 5 vivid by imagining that, given MWI, only Up branches are life-permitting. This ensures that if both Up and Down branches exist, you observe the Up branch. So if there is an Up branch, you observe it. (If ST is true and the spin collapses to Up, you observe Up; if ST is true and the spin collapses to Down, there is no Up branch and you observe Down.) The evidence is now:

10  $E_c^B =$  I learn that Up occurs in this branch by a procedure biased towards Up branches.

As the procedure is biased towards Up branches,  $P(E_c^B | MWI) = 1$ . Recall that  $P(E_c | ST) = 1/2$  for any procedure, so  $P(E_c^B | MWI) > P(E_c^B | ST)$ . Thus, the  
 15 evidence would confirm MWI, just as with Naïve Conditionalization. So it looks like Naïve Conditionalization is similar to a maximally biased procedure.

But there is an important difference. In our analysis, MWI is confirmed by the procedure that is biased towards Up successfully selecting an Up branch. If instead it selects a Down branch, then MWI is eliminated (ST and Down  
 20 would be the only possibility left). There is no automatic confirmation of MWI.

But Naïve Conditionalization says there is automatic confirmation of MWI, whatever the evidence. It says that an observation of Up confirms MWI and an observation of Down confirms MWI. This requires that the procedure is  
 25 biased towards Up if Up is observed, and biased towards Down if Down is observed. So Naïve Conditionalization must be rejected; even if it is defensible to argue that there is a bias in one direction, there cannot be a bias in both directions for an in-branch observer. The in-branch observer can only observe the branch she is on, so properties in the other branch will not be observed  
 30 by her.

Suppose instead that there is no bias. That is, we have a random procedure—Up branches have the same probability of being observed as Down branches, so you are just as likely to be in an Up branch as a Down branch. So the new evidence is:

35  $E_c^R =$  I learn that Up occurs in this branch by a random procedure.

you are necessarily the inhabitant of the Up branch, this is not a priori, nor is it certain, so your credence will be less than 1.

As the procedure is random,  $P(E_c^R | MWI) = 1/2$ ;  $P(E_c^R | ST) = 1/2$  as before, so MWI is not confirmed. Thus we avoid the absurd result that MWI is confirmed by any observation.

5 The main aim of this article is to highlight the role of selection effects, and this is what I hope to have just done regarding MWI. But the secondary aim is to defend the selection effect that I think correctly models the case. In MWI, I think the random procedure correctly models the way we acquire our evidence. To assume instead that we are more likely to observe an Up branch is implausible and unmotivated—Up branches are just as hospitable to life, and just as likely to be observed as Down branches.<sup>14</sup> I will defend this in  
10 Section 6.

#### 4 Sleeping Beauty

I will now offer a parallel analysis of Sleeping Beauty (Elga [2000]). The Sleeping Beauty problem runs as follows:

15 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  
20 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.

25 Note that Heads and Tails are two possible worlds and the days are centres within those worlds.

	Monday	Tuesday
Heads	Awake	
Tails	Awake	Awake

30 When she is woken, what credence should she have that the coin landed Heads? Some say that her credence in Heads should stay at  $1/2$ . Call these *Halfers* (Lewis [2001]). Some say that her credence in Heads should fall to  $1/3$ . Call these *Thirders* (Elga [2000]).  
35

<sup>14</sup> This is not to say that every Up branch is as likely to be observed as every Down branch. The branches might have unequal weightings. This will be matched by unequal probabilities of Up and Down in ST. MWI will only be confirmed by observing Up if the probability that this branch is Up given MWI is greater than the probability that this branch is Up given ST. This would require a selection effect favouring Up that is not reducible to the branch weight. Experiments have not shown such favouring. Nevertheless, those worried by unequal weightings can regard my arguments as applying only to the simple case in which there are equal weights.

A plausible sounding argument (which I will reject) for Thirdering runs as follows:

The Technicolour Beauty Argument

5 The difficulty in the Sleeping Beauty problem comes from the fact that Beauty has no expression that allows her to refer uniquely to the different days. Let's allow her to refer uniquely to the different days by making them subjectively distinguishable. Suppose that she sees a coloured piece of paper that is either red or blue on each day. If she is woken once (Heads), she sees either a red or blue piece of paper with probability 0.5. If she is woken twice  
10 (Tails), she sees a red piece of paper on one day and a blue piece of paper on the other. Beauty knows all this. Suppose Beauty observes the red piece of paper.

$E$  = There is a waking on which the red paper is observed.

15 (The colour of the paper on any given day is independent of the coin flip, so the additional information regarding the coloured paper doesn't affect what Beauty should believe about the coin flip. So Beauty's credences in this case should match her credences in the original Sleeping Beauty.)

Does  $E$  confirm Tails? This amounts to: is  $P(E|Tails) > P(E|Heads)$ ?  
20 Yes. Tails says that both the red and blue paper will be observed, so  $P(E|Tails) = 1$ , while  $P(E|Heads)$  is  $1/2$ . It follows from Bayes' theorem that  $P(Heads|E) = 1/3$ , just as Thirder's claim.

25 Technicolour Beauty

	<b>Monday</b>	<b>Tuesday</b>
<b>Heads</b>	Red or Blue	
<b>Tails</b>	Red or Blue	Red or Blue (whichever wasn't seen on Monday)

30 Notice that it didn't matter that red paper was observed rather than blue paper. Both outcomes had a probability of 1 given Tails, both outcomes had a probability of  $1/2$  given Heads, so both confirm Tails. We get the result that any possible evidence confirms Tails.

The argument above is based on (Titelbaum [2008]), who identifies the  
35 problem in Sleeping Beauty as the inability to express on Sunday the evidence learnt on Monday. He fixes this problem by allowing that the awakenings are distinguishable, while not allowing that this distinguishability helps Beauty figure out which day it is. (I should add that Titelbaum introduces Technicolour Beauty within a more general formal framework, and does not express the argument as I have done above. Nevertheless, he uses the  
40 likelihoods  $P(E|Tails) > P(E|Heads)$  to generate his Thirder position. (See *ibid.*, p. 593, Table 10, assumption 10 and Table 11 assumption 6). I don't dispute the values. I dispute that  $E$  expresses Beauty's total evidence.)

The kind of thinking in Technicolour Beauty also seems to be endorsed by Thirder Papineau and Durà-Vilà ([2009a], [2009b]) in their dispute with Peter Lewis ([2007], [2009]). Lewis had argued that Sleeping Beauty and the Everett interpretation were sufficiently analogous that Everettians had to be committed to the Halfer answer in Sleeping Beauty. (I will argue that he is right.) In both cases, the two hypotheses are that there is one location of observation versus two locations of observation. And in both cases, the agent doesn't know which location she is at. To use Lewis's apt phrase, she gets lost in the branches.

	Quantum Mechanics		Sleeping Beauty		
	<b>Branch 1</b>	<b>Branch 2</b>		<b>Monday</b>	<b>Tuesday</b>
<b>ST</b>	Up or Down		<b>Heads</b>	Awake	
<b>MWI</b>	Up	Down	<b>Tails</b>	Awake	Awake

Adding the coloured paper and removing the names of the days makes the analogy even more transparent:

	Quantum Mechanics		Technicolour Beauty		
	<b>Branch 1</b>	<b>Branch 2</b>		<b>Day 1</b>	<b>Day 2</b>
<b>ST</b>	Up or Down		<b>Heads</b>	Red or Blue	
<b>MWI</b>	Up	Down	<b>Tails</b>	Red	Blue

Even though both Up and Down are certain to happen given MWI, quantum theory requires non-trivial probabilities regarding which branch is observed, so Everettians have gone to great lengths to devise ways of assigning non-trivial probabilities to Up and Down (Saunders [1998]; Vaidman [2002]; Greaves [2004]; Papineau [2004]; Wallace [2006]). Peter Lewis argued that if there really are non-trivial probabilities in MWI regarding which branch is observed, then by analogy there should be non-trivial probabilities in Tails (in Sleeping Beauty) regarding which day is observed. If so, we could model Tails as a case in which a second fair coin is flipped to determine whether Beauty will be woken on Monday or Tuesday. If so, the Halfer answer is clearly correct.

But Papineau and Durà-Vilà argue that although the outcomes have non-trivial probabilities in MWI, there is only trivial probability (i.e. probability 1) with Tails in Sleeping Beauty. Why? Because credences should match objective probabilities and 'the objective probabilities of waking on Monday and Tuesday are both 1'<sup>15,16</sup> ([2009b], p. 90). This mirrors the Technicolour

<sup>15</sup> Due to the brevity of their discussion, I'm not certain Papineau and Durà-Vilà would endorse the Technicolour Beauty argument. But it does seem like a natural reading of their papers.

<sup>16</sup> If the appropriate probability were 1, it's not clear to me what would stop an opponent from responding that by analogy the appropriate probability in MWI is also 1. Papineau and Durà-Vilà base their argument on the 'objective quantum probabilities'. But the status of these 'objective quantum probabilities' and their connection to rational credence is one of the issues up for debate.

Beauty argument that the probabilities of there being a red paper and a blue paper are both 1 given Tails;  $P(E|Tails) = 1$ , while  $P(E|Heads) = 1/2$ , and Thirring follows.

I will now argue that the Technicolour Beauty argument fails to take into account the selection effects, and so assigns the wrong conditional probabilities.<sup>17</sup> Titelbaum and Papineau & Durà-Vilà seem to take  $E$  (i.e., There is a waking on which the red paper is observed), or something like it, to express the evidence. But the total evidence is stronger than  $E$ . The evidence learnt when red paper is observed is not the uncentred proposition that there is an awakening on which the red paper is seen, but the centred proposition that red paper is observed on *this* awakening. The agent has learnt when it is—it is a day with red paper. I will argue that this has probability 1/2, rather than 1. The first thing we need to do is ensure that the evidence is a centred proposition:

$E_c =$  Today is a red paper day.

The second thing we need to do is put in the correct selection effect. If Heads lands, the selection effect is trivial. As she is only awake on Monday, she will see whatever coloured paper is placed before her on Monday, so  $(E_c|Heads) = 1/2$ . However, in order to determine  $P(E_c|Tails)$ , we have to ask: given that Beauty is woken twice, what should her subjective probability be that her current time-slice is on a day on which red paper is observed?<sup>18</sup> The answer to this question will depend on whether there is any bias in the selection procedure.

Suppose first that the procedure is biased towards red paper days. Recall this means that if there is a red paper day then red paper is observed. To make this vivid, modify the case so that if Tails lands, Beauty is woken only on the red paper day, whether it be Monday or Tuesday. (If Heads lands, she is woken only on Monday as before, and sees whatever paper is placed before her on Monday.) So the new evidence is:

$E_c^B =$  I learn that today is a red paper day by a procedure biased towards red paper days.

As the procedure is biased towards red paper days,  $P(E_c^B|Tails) = 1$ . As  $P(E_c^B|Heads) = 1/2$ , Tails is confirmed.

This might appear to give us the confirmation of Tails defended by Titelbaum and Papineau & Dura-Vila. But they think Tails is confirmed *whatever* observation is made, and this isn't possible. Tails is confirmed if

<sup>17</sup> I emphasize that I am concerned with the conditional probabilities on a day the paper is observed. It is a further question how these relate to the Sunday probabilities, which I take no stand on in this paper. See footnote 10.

<sup>18</sup> Again, it doesn't matter if the awakening is individuated by the day. We are using subjective probability, so what matters are Beauty's *beliefs* about what day it is. See footnote 12.

the property observed is the property towards which the procedure is biased (red paper days in this case). If instead a procedure which is biased towards red had selected a blue paper day, Tails would be eliminated (Heads and blue paper would be the only possibility left). The view that both observations confirm Tails requires that the procedure is biased in both ways at once. And this isn't possible for an agent who can only remember the current day. Just as Naïve Conditionalization failed because an in-branch agent can't be modelled using a procedure biased towards both Up and Down, automatic confirmation of Tails fails because a within-day agent can't be modelled using a procedure biased towards both red and blue paper days.

Suppose instead that there is no such bias towards observing a red paper day. Then we have a random procedure—red paper days have the same probability of being observed as blue paper days.

$E_c^R = I$  learn that today is a red paper day by a random procedure.

As the procedure is random,  $P(E_c^R | \text{Tails}) = 1/2$ . As  $P(E_c^R | \text{Heads}) = 1/2$ , Tails is not confirmed.

My main concern is simply to highlight the importance of selection effects. But I also want to defend the answers I favour. I think in this case it is even more plausible than in MWI that the process should be modelled by a random procedure. Indeed, we stipulated that given Tails, Beauty would see a red paper one day and a blue paper the next. The only way the procedure would not be random would be if, on one of the awakenings before seeing the paper, Beauty had a reasonable subjective degree of belief that 'today is a red paper day given Tails' that diverged from 50%. Knowing the setup as she does, it is hard to see how Beauty could reasonably believe such a thing. Nevertheless, can we give a positive argument that Beauty should assign 50% credence to 'this is a red paper day given Tails'? I will argue in the final section on the restricted principle of indifference that we can. But first I want to clarify my position and its significance.

## 5 Significance of the Analogy

I have argued that Sleeping Beauty has the same structure as MWI, shown most clearly by this table, repeated from above:

	Quantum Mechanics			Technicolour Beauty	
	Branch 1	Branch 2		Day 1	Day 2
ST	Up or Down		Heads	Red or Blue	
MWI	Up	Down	Tails	Red	Blue

What follows? Peter Lewis ([2007], [2009]) argues that Everettians must be Halfers. I agree. He thinks this is bad news for Everettians. I think it is good

news for Everettians, but I won't debate the issue here, as the significance of the analogy runs deeper. The issue isn't about whether we accept MWI, but whether we gain suspiciously easy evidence for MWI.

5 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 regarding MWI is that we gain some reason after branching to believe that the possible world in which there are more observations (MWI) is more probable than the one with fewer (ST). As branching is happening all  
10 the time, it would follow that we have overwhelming evidence in favour of MWI!

For example, suppose I see it is raining, where some quantum probability is responsible for it raining rather than being sunny. According to MWI, it was certain that rain would be observed (in some branch), but according to ST it is  
15 less than certain. So our everyday observations are constantly confirming MWI. On this reasoning, MWI gets enormous confirmation without the need for modern physics. The Ancients could have worked out that they have overwhelming evidence for MWI merely by realizing it was a logical possibility and observing the weather.

20 If Thirders are to reject this easy evidence for MWI, they owe us an explanation of where the disanalogy lies between Sleeping Beauty and MWI. This isn't just a problem for Thirders who are Everettians, it's a problem for any Thirders who think that evidence for MWI should not be easy to come by.<sup>19</sup>

So even those who dismiss the scientific theory that the universe is dividing  
25 face the question of what would happen to confirmation theory if it were. The fact that philosophers have been led to some of these problems by science is accidental. Interest in personal identity under cases of fission, for example, was not motivated by the invention of fission machines, but by interest in what fission cases tell us about our concept of personal identity. Similarly, fission of  
30 the world is not only interesting because a scientific theory says that the world really is dividing, but because of what fission cases tell us about our concept of confirmation.

The final loose end is to defend the assumption that observers in Tails and MWI should assign a probability of 50% to being each of the two observers.  
35 This was necessary to model the cases with a random procedure. Adam Elga

<sup>19</sup> Those who accept that evidence for MWI is easy to come by might endorse the Self-Indication Assumption (Bostrom [2002]), which provides a priori confirmation of hypotheses with greater populations. Bostrom rejects it, but similar positions are defended by Bartha and Hitchcock ([1999]), Dieks ([2007]) and Peter Lewis ([2010]), who all claim that it cancels out the shift in favour of smaller populations of the Doomsday Argument. But it is not clear that there is an analogous cancellable shift with regard to MWI.

([2004]) has defended a restricted principle of indifference<sup>20</sup> which defends this assumption. Indeed, the intuitive case Elga uses to motivate his principle is exactly the same as Tails in Sleeping Beauty. I will explain Elga's arguments and defend them from the criticisms of Weatherson ([2005])

## 6 The Restricted Principle of Indifference

I include this section against my better judgment, as it may give the impression that my earlier arguments are based on those in this section. But my main point has been that Naïve Conditionalization and the analogous thirder arguments in Sleeping Beauty are incorrect. Those arguments rely on assumptions about selection effects that are not just false but incoherent - they require that the procedures are biased in both directions. Even if the procedures are biased in one direction, the arguments for Naïve Conditionalization and being a Thirder still fail. I do not need to argue for the further point that the procedures are in fact random. Nevertheless, in this section I will do exactly that. Elga motivates his argument with the following scenario—O'Leary is locked in the boot of his car all night, knows he will wake at 1 a.m. and 2 a.m., but both awakenings will be subjectively indistinguishable because by 1 a.m. he will have forgotten the 2 a.m. awakening. In both the Tails and O'Leary scenarios, it is highly plausible that the agent's credence at each centre should be 1/2. More precisely, Elga argues that centred worlds deserve equal credence when:

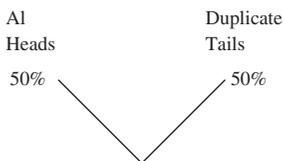
- (a) they are in the same uncentred world; and
- (b) the agents are subjectively indistinguishable.

Elga's argument is based on the following thought experiment. Suppose an agent named Al is in a lab and about to be drugged. If a fair coin lands Heads, he will be woken up unharmed. But if the coin lands Tails, the scientists will make an exact Duplicate of Al and put the original in a coma. Al and the Duplicate will be unable to tell when woken which agent he is. Al knows all this. Before being drugged, Al's credence in Heads should be 50%, as the Principal Principle dictates (Lewis [1980]). Elga argues that after waking, Al's credence in Heads should still be 50%. The reason is that Al has 'neither gained nor lost information relevant to the toss outcome'<sup>21</sup> (Elga, *ibid.*,

<sup>20</sup> Do not confuse Elga's restricted principle of indifference with a less restricted principle of indifference that would imply equal credence in each awakening: Monday & Heads, Monday & Tails, Tuesday & Tails. As David Lewis ([2001], p. 172, fn 1) says 'This would afford a swift shortcut to Elga's conclusion—much too swift—and Elga is wise to have nothing to do with it.'

<sup>21</sup> This is an odd thing for Elga to rely on, as he seems to explicitly reject this premise in his earlier paper defending the Thirder position. There, he concludes that 'the manner in which an agent counts her own temporal location as relevant to the truth of some proposition can change over time [...] even during a period in which that agent neither receives new information nor suffers a cognitive mishap' ([2000], p. 146). I am grateful to an anonymous referee for stressing this point.

p. 392). If this is right, then the restricted principle of indifference follows with fairly minimal assumptions (I refer the reader to Elga's paper for the details).



Weatherson ([2005]) disagrees with Elga, and argues that Al's degree of belief in Heads should not necessarily remain the same after waking. Weatherson offers four considerations in support of his view. I will very briefly go through them and indicate why I am not convinced.

Firstly, Weatherson claims that Al gets some evidence on waking—perhaps he experiences certain sensations or has certain thoughts—and this evidence might be relevant to Heads, for all he knows.

But Elga set up the thought experiment so that Al and the Duplicate are subjectively indistinguishable. This means that any sensations or thoughts Al has would also be experienced or thought by the Duplicate, so they cannot be relevant to Heads.

Second, Weatherson charges that Elga is committed to internalism about evidence, which conflicts with the evidence externalism of Williamson ([1998]) and Campbell ([2002]). Moreover, it conflicts with the (epistemic) possibility that such evidence externalism *might* be true.

But I think we can grant Weatherson the controversial thesis that evidence externalism is, or might be, true. Elga doesn't need to say that Al and the Duplicate would have the same *evidence*. He just needs to say that Al and the Duplicate would be in the same subjective state, and that one's credences should depend on one's subjective state. As Al and the Duplicate would be in the same subjective state, no part of this subjective state can support Heads.

Third, Weatherson charges that Al's degree of belief in Heads should change from being risky to being uncertain after the toss. That is, it should be represented by an interval rather than a point probability. (The distinction between risk and uncertainty is due to Knight [1921]; the central point is that risk is measurable, and uncertainty immeasurable.)

But it would be odd for a risky proposition to become uncertain in this way. Uncertainty is most plausible in cases in which we have very little information regarding the proposition. But in this case we have as much information regarding Heads after waking as we did before—we know that it is a fair coin. Furthermore, Elga ([2010]) has put forward an argument that holding interval probabilities is irrational that I at least find persuasive.<sup>22</sup>

<sup>22</sup> The literature on dilation (see Seidenfeld and Wasserman [1993]) might help Weatherson here, but he doesn't mention it, and a discussion would take us too far afield.

Fourth, and relatedly, even if there is no reason for Al to change his credence in Heads in one direction or the other after the toss, Weatherson suggests it may still be reasonable for his credences to change. The absence of a reason to change is not a reason to stay put.

5 But Al already does have a reason to stay put—his credence in Heads matches the objective chances, and it is widely accepted that credences should match known objective chances. Without some defeater for this reason, surely he should stay put.

10 Obviously, this has been the quickest of discussions, and there is much more to be said. Even if Elga's specific argument can be faulted however, some such restricted principle of indifference seems highly plausible. Scientists continue to use principles of indifference to great effect (Jaynes [1973]), even if philosophers have not produced a satisfactory justification. As MWI is a scientific theory, it does not seem objectionable to use a principle of  
15 indifference, even if we were to lack a solid philosophical justification for it at present.

Both the Many Worlds Interpretation and Sleeping Beauty involve learning self-locating evidence. I have argued that the selection procedure by which the self-locating evidence was discovered must be taken into account.  
20 In particular, I have argued that when we do so, we find that the observation of the result of a measurement does not automatically confirm the Many Worlds interpretation. A parallel analysis can also be applied to the Sleeping Beauty problem. The result is that a certain argument for the Thirder position is shown to be problematic. Finally, the analogy between Sleeping Beauty and the Many Worlds Interpretation means that  
25 Thirder appear to have suspiciously easy evidence that the Many Worlds Interpretation is true.

### Acknowledgements

30 Thanks to Victor Durà-Vilà, Kenny Easwaran, Adam Elga, Branden Fitelson, Hilary Greaves, Peter Lewis, Alan Hájek, David Papineau, John Perry, Huw Price, Mike Titelbaum, David Wallace, and Brian Weatherson, audiences at the Coffs Harbour Rationality Conference and the University of Sydney and two referees for helpful discussion and comments.

35 *160 Convent Ave.  
New York  
NY 10031  
USA  
bradleydarren@gmail.com*

## References

- Bartha, P. and Hitchcock, C. [1999]: ‘No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes’s Theorem’, *Philosophy of Science*, **66**. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association, Part I: Contributed Papers, pp. S339–53.
- 5 Bostrom, N. [2002]: *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, New York: Routledge.
- Bovens, L. and Ferreira, J. L. [2010]: ‘Monty Hall drives a wedge between Judy Benjamin and Sleeping Beauty: a Reply to Bovens’, *Analysis*, **70**, pp. 473–81.
- 10 Bradley, D. J. [2009]: ‘Multiple Universes and Observation Selection Effects’, *American Philosophical Quarterly*, **46**, pp. 61–72.
- Bradley, D. J. [2010]: ‘Conditionalization and Belief De Se’, *Dialectica*, **64**, pp. 247–50.
- Bradley, D. J. [forthcoming]: ‘Self-location is no problem for conditionalization’, forthcoming in *Synthese*, <dx.doi.org/10.1007/s11229-010-9748-9>.
- 15 Campbell, J. [2002]: *Reference and Consciousness*, Oxford: Oxford University Press.
- Dieks, D. [2007]: ‘Reasoning about the Future: Doom and Beauty’, *Synthese*, **156**, pp. 427–39.
- Eddington, A. [1939]: *The Philosophy of Physical Science*, Cambridge: Cambridge University Press.
- 20 Elga, A. [2000]: ‘Self-Locating Belief and the Sleeping Beauty Problem’, *Analysis*, **60**, pp. 143–7.
- Elga, A. [2004]: ‘Defeating Dr. Evil with Self-Locating Belief’, *Philosophy and Phenomenological Research*, **69**, pp. 383–96.
- 25 Elga, A. [2010]: ‘Subjective Probabilities Should be Sharp’, *Philosopher’s Imprint*, **10**, <hdl.handle.net/2027/spo.3521354.0010.005>.
- Freund, J. [1965]: ‘Puzzle or Paradox?’, *American Statistician*, **19**, pp. 29–44.
- Glymour, C. [1980]: *Theory and Evidence*, Princeton: Princeton University Press.
- Godfrey-Smith, P. [2003]: ‘Goodman’s Problem and Scientific Methodology’, *The Journal of Philosophy*, **100**, pp. 573–90.
- 30 Greaves, H. [2004]: ‘Understanding Deutsch’s Probability in a Deterministic Universe’, *Studies in History and Philosophy of Modern Physics*, **35**, pp. 423–56.
- Greaves, H. [2007]: ‘Probability in the Everett Interpretation: A Solution to the Epistemic Problem’, *Studies in History and Philosophy of Modern Physics*, **38**, pp. 120–52.
- 35 Groisman, B. [2008]: ‘The End of Sleeping Beauty’s Nightmare’, *British Journal for the Philosophy of Science*, **59**, pp. 409–16.
- Hacking, I. [1965]: *The Logic of Statistical Inference*, Cambridge: Cambridge University Press.
- 40 Hacking, I. [1967]: ‘Slightly More Realistic Personal Probability’, *Philosophy of Science*, **34**, pp. 311–25.
- Horwich, P. [1982]: *Probability and Evidence*, Cambridge: Cambridge University Press.
- Howson, C. and Urbach, P. [2006]: *Scientific Reasoning: The Bayesian Approach*, 3rd edition; Chicago: Open Court.

- Hutchison, K. [1999]: ‘What Are Conditional Probabilities Conditional On?’, *British Journal for Philosophy of Science*, **50**, pp. 665–95.
- Jackson, F. [1975]: ‘Grue’, *The Journal of Philosophy*, **72**, pp. 113–31.
- Jackson, F. and Pargetter, R. [1980]: ‘Confirmation and the Nomological’, *The Canadian Journal of Philosophy*, **10**, pp. 415–28.
- Jaynes, E. T. [1973]: ‘The Well-Posed Problem’, *Foundations of Physics*, **3**, pp. 477–92.
- Knight, F. H. [1921]: *Risk, Uncertainty and Profit*, Chicago: Houghton Mifflin Company.
- Kotzen, M. [unpublished]: ‘Multiple Studies and Evidential Defeat’, <philosophy.fas.nyu.edu/docs/IO/1328/sample.pdf>.
- Leslie, J. [1989]: *Universes*, London: Routledge.
- Lewis, D. K. [1979]: ‘Attitudes *De dicto* and *De se*’, *The Philosophical Review*, **88**, pp. 513–43.
- Lewis, D. K. [1980]: ‘A Subjectivist’s Guide to Objective Chance’, in R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Volume II., Berkeley: University of California Press, pp. 263–93.
- Lewis, D. K. [2001]: ‘Sleeping Beauty: Reply to Elga’, *Analysis*, **61**, pp. 171–6.
- Lewis, P. J. [2007]: ‘Quantum Sleeping Beauty’, *Analysis*, **67**, pp. 59–65.
- Lewis, P. J. [2009]: ‘Reply to Papineau and Durà-Vilà’, *Analysis*, **69**, pp. 86–9.
- Lewis, P. J. [2010]: ‘A Note on the Doomsday Argument’, *Analysis*, **70**, pp. 27–30.
- Meacham, C. [2008]: ‘Sleeping Beauty and the Dynamics of *De se* Belief’, *Philosophical Studies*, **138**, pp. 245–69.
- Papineau, D. [2004]: ‘David Lewis and Schrödinger’s Cat’, *Australasian Journal of Philosophy*, **82**, pp. 153–69.
- Papineau, D. and Durà-Vilà, V. [2009a]: ‘A Thirder and an Everettian: a Reply to Lewis’, *Analysis*, **69**, pp. 78–86.
- Papineau, D. and Durà-Vilà, V. [2009b]: ‘Reply to Lewis: Metaphysics Versus Epistemology’, *Analysis*, **69**, pp. 89–91.
- Perry, J. [1979]: ‘The Problem of the Essential Indexical’, *Nous*, **13**, pp. 3–21.
- Price, H. [unpublished]: ‘Probability in the Everett World: Comments on Wallace and Greaves’, <philsci-archive.pitt.edu/archive/00002719/>.
- Salmon, W. C. [1975]: ‘Confirmation and Relevance’, in G. Maxwell and R. M. Anderson, Jr. (eds), *Minnesota Studies in the Philosophy of Science, Volume 6, Induction, Probability, and Confirmation*, Minneapolis: University of Minnesota, pp. 3–36.
- Saunders, S. [1998]: ‘Time, Quantum Mechanics and Probability’, *Synthese*, **114**, pp. 373–404.
- Schwarz, W. [unpublished]: ‘Changing Minds in a Changing World’, <www.umsu.de/words/belief.pdf>.
- Seidenfeld, T. and Wasserman, L. [1993]: ‘Dilation for Sets of Probabilities’, *The Annals of Statistics*, **21**, pp. 1139–54.
- Shafer, G. [1985]: ‘Conditional Probability’, *International Statistical Review*, **53**, pp. 261–77.
- Titelbaum, M. G. [2008]: ‘The Relevance of Self-Locating Beliefs’, *Philosophical Review*, **117**, pp. 555–606.

- Vaidman, L. [2002]: 'The Many-Worlds Interpretation of Quantum Mechanics', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, <<http://plato.stanford.edu/archives/sum2002/entries/qm-manyworlds/>>.
- Vos Savant, M. [1997]: *The Power of Logical Thinking*, New York: St Martin's Griffin.
- 5 Wallace, D. [2006]: 'Epistemology Quantized: Circumstances in which we should come to Believe in the Everett Interpretation', *British Journal for the Philosophy of Science*, **57**, pp. 655–89.
- Weatherson, B. [2005]: 'Should We Respond to Evil with Indifference?', *Philosophy and Phenomenological Research*, **70**, pp. 613–35.
- 10 Williamson, T. [1998]: 'Conditionalizing on Knowledge', *British Journal for the Philosophy of Science*, **49**, pp. 89–121.