Four Problems About Self-Locating Belief

Introduction

I will argue that four problems that appear to be very different have the same structure. I give a unified treatment of the Doomsday Argument, Sleeping Beauty, the Fine-tuning Argument and the Everett interpretation of quantum mechanics. My analysis uses simple models of the cases and no controversial assumptions about confirmation or self-locating evidence. I will argue that the troublesome feature of all these cases is not self-location, but observation selection effects.

In part 1 I explain observation selection effects and develop a framework for analyzing them using some toy examples. In part 2 I explain the four problems about self-locating belief and analyze them using our framework. In part 3 I criticize other approaches to self-locating beliefs by drawing on the analogy between the four problems.

Part 1: Selection Effects and Self-Location

1.1 Selection Effects

Whenever a sample is drawn from a population, some particular method must be used. Call this method the selection procedure. The effect the selection procedure has on the inference is the selection effect. Eddington's (1939) classic example involves fishing with a net[1]. 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 procedure, but we will only need two:

A *random* procedure is one where each member of the population has an equal chance of being selected for the sample. This is familiar from statistics (Stuart 1962). We will assume that random procedures always select at least one object.

---

[1] Selection effects play an important role in Horwich's 1982 analysis of the ravens paradox and most famously in the Monty Hall problem (vos Savant 1997). See Hutchison 1999 and Bostrom 2002 for systematic discussions, and also Kotzen (forthcoming).

Of the many non-random procedures, we will need only the following:

A *biased* procedure is one such that given that there is an object with property p in the population, an object with p is selected for the sample. We'll say the procedure is biased towards p in such a case. Formally:
A procedure is biased towards p iff Pr(an object with p is selected | p is instantiated in the population) = 1.[2] (Careful: This is different from the ordinary language use of 'bias'; my use is closer to the ordinary language 'maximal bias'.)

To get a grip on these selection procedures, let's consider two simple thought-experiments that have the same structure as the four cases we will look at. We'll assume throughout that the agent has certain knowledge of the procedure.

1.2 Uncertain Small Ball

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.)

	Ball 1	**Ball 2**
Heads	Large or small	
Tails	Large	Small

H = Heads
T = Tails

---

[2] This procedure is neutral on what happens if there is no object with p in the population, or exactly which objects are selected if there is more than one object with p in the population. 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.

A sample of one is taken from the urn. It turns out to be small.
How should we describe the evidence? One thing that has been learnt is:
E = There is a ball which is small.
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. Let's run through the effects of knowing there is a random or a biased procedure.

Random Procedure
First suppose the procedure is random. That is, every ball in the urn has an equal chance of being selected. We could make this vivid by imagining that a large hole is opened and the urn shaken until a ball comes out.

E_R = I learn 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% chance of the urn containing a small ball, so $P(E_R|H) = 1/2$. Assuming the standard view that E confirms H iff $P(E|H) > P(E|-H)$ (Salmon 1975) neither Heads nor Tails is confirmed. Discovering a property with a random procedure confirms neither hypothesis if the probability that a randomly selected object has that property is the same given either one ball or two. (We'll see something similar happens in the Everett interpretation of quantum mechanics.)

Biased Procedure
Now assume the procedure is biased towards smallness. That is, given that a small ball is in the urn, a small ball is selected. 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.

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

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. That is, $P(E_B|H) = 1/2$ and $P(E_B|T) = 1$. $P(E_B|H) > P(E_B|T)$, so E_B confirms Tails. Discovering, with a biased procedure, a property that is more likely to be instantiated given two balls than one, confirms the two ball hypothesis. (We'll see something similar happens in the Multiverse case.)

Notice that in Uncertain Small Ball, the property discovered (being a small ball) might not be instantiated – this would happen if Heads lands and a big ball is in the urn. The results are summarized in table 1.

	p might not be instantiated
Random procedure	No confirmation

Biased procedure	Two Outcomes confirmed

Table 1

1.3 Certain Small Ball

Let's now suppose that the property discovered (being a small ball) *is* certain to be instantiated. We can do this by supposing that a small ball is always placed in the urn if Heads lands. Everything else is as before.

	Ball 1	**Ball 2**
Heads	Small	-
Tails	Small	Large

This changes the inferences we can draw when a small ball is selected by either procedure.

Random Procedure

Assume a large hole is opened so each ball has an equal chance of being selected. If Heads, the probability of a small ball being selected is 1, as there is only the one small ball in the urn (here we invoke the assumption that at least one object is selected). If Tails, the probability of a small ball being selected is 1/2. $P(E_R|H) > P(E_R|T)$, so Heads is confirmed. Discovering, by a random procedure, a property that is certain to be instantiated confirms that there is only one ball if the second ball has some chance of not having the property and being selected. This is because the presence of such a second ball would lower the probability that a small ball would be selected to less than 1. (We'll see something similar happens in the Doomsday Argument.)

Biased Procedure

Assume a small hole is opened and the urn shaken, so the procedure is biased towards smallness. There is certain to be a small ball in the urn (given Heads or Tails) and the small ball is certain to be selected, so $P(E_B|H) = P(E_B|T) = 1$, so neither Heads nor Tails is confirmed. Think of this as the biased procedure searching for a property that is certain to be instantiated. Discovering, by a biased procedure, a property that is certain to be instantiated confirms neither hypothesis. Notice that even if we then add more objects to the Heads hypothesis, there will still be no confirmation. (We'll see something similar happens in Sleeping Beauty.)

These results are added to Table 1 to get Table 2.

	p is certain to be instantiated	**p might not be instantiated**
Random procedure	One Outcome confirmed	No confirmation
Biased procedure	No confirmation	Two Outcomes confirmed

Table 2

Table 2 is a map of this paper. I will argue that each of the four problems of self-locating belief fits into a different box. Let's now introduce self-locating beliefs and connect them to selection effects.

1.4 Self-locating Belief

Traditionally, confirmation theorists worked with eternal beliefs that always have the same truth value. But consider self-locating beliefs which locate the agent at a certain time or location (or possibly with respect to other variables, as we'll see). What if the evidence is a self-locating belief? Two issues are raised.

The first is that self-locating beliefs change in ways that non-self-locating beliefs don't. This is because self-locating beliefs 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 relatively uncontroversial that the Bayesian framework – conditionalization in particular – must be modified to incorporate this kind of belief change[3]. Various theories have been proposed (Meacham 2005, Titelbaum 2008, Schwarz forthcoming) but I won't say much about this kind of belief change here (I discuss this in Bradley forthcoming a and forthcoming b).

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. 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 belief. Instead, something has been discovered that was previously uncertain.

---

[3] Though see Stalnaker (2010) for dissent.

The authors mentioned two paragraphs above 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 the familiar view that E confirms H iff P(E|H) > P(E|-H).

Instead, the complications in such cases are caused by the fact that E might be true without being learnt. And this feature provides the connection to selection effects. Although you select a small ball from a population, the small ball could have been in the population without being observed (a large ball could have been selected instead). Similarly, although you look at your watch at 12:30, it could have been 12:30 without you observing it (you could have been asleep).

By contrast, in many familiar examples – rolling a dice, drawing a card, reading a measuring device – it is natural to assume that if E is true then E is learnt. As a result, the assumption (that if E, then E is learnt) is usually made without fanfare and enforced thereafter. For example, Hacking (1967 p. 316, 324) treats the inference from the truth of E to learning E as a 'trifling idealization' and Howson and Urbach (2006 p. 54) makes use of an 'omniscient oracle'. But this is a substantive assumption, which we are modelling as a biased procedure.

There is also a disanalogy between sampling and self-location cases that is worth pointing out. Talk of selection procedures is literal in sampling cases but metaphorical in self-location cases, where selection procedures are just a way of making vivid the possible gap between E and learning E. Whereas a literal procedure generates physical probabilities (objective chances), our metaphorical procedures generate subjective (or inductive) probabilities (Maher 2006). So what matters for us is what the agent believes (or ought to believe) about the probability of learning the evidence given the hypotheses. I will mostly leave reference to the agent's beliefs implicit and simply speak of the procedure we should use to model the case.

Part 2: Four Problems

All four problems of self-locating belief have the following structure:

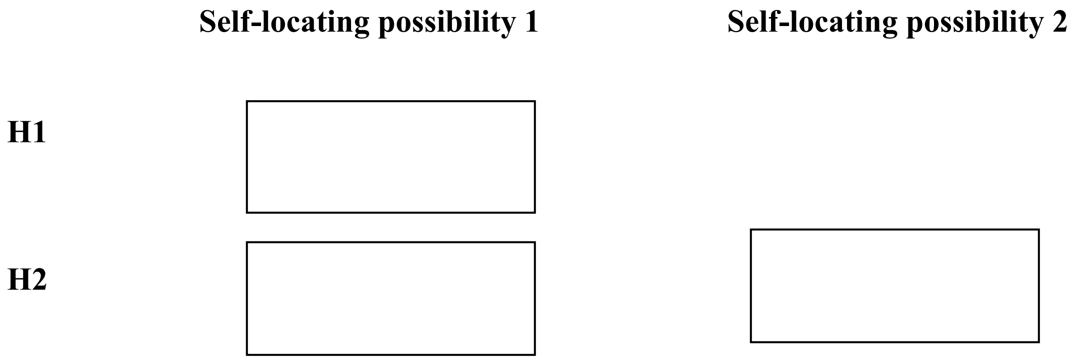

Figure 1: Self-locating possibilities

There are two possible worlds, with two possible locations in each world. Each of the locations has a particular property. The question is whether learning that you are at one of the locations, with some particular property, confirms either hypothesis. We will see that the answer depends on two variables (see table 3) –the procedure by which the property was discovered (rows) and whether the property discovered was certain to be instantiated (columns). I will argue that the four problems about self-locating belief fit into the four boxes as follows.

	p is certain to be instantiated	**p might not be instantiated**
Random procedure	One Outcome confirmed Doomsday Argument	No confirmation Quantum mechanics
Biased procedure	No confirmation Sleeping Beauty	Two Outcomes confirmed Fine-Tuning

Table 3

2. 1 Everettian Interpretation of Quantum Mechanics

Stochastic quantum mechanics (QMS) is an indeterminstic theory. It says that when a measurement is made, there is a certain chance of each possible outcome occurring. Everettian quantum mechanics (QME) is a deterministic theory. It says that when a measurement is made, the universe divides, and there is a branch in which each outcome occurs. If we don't know which of these is true, and we are about to measure the spin of a particle in a nonextreme superposition of Up and Down, we face the following possibilities:

	Branch 1[4]	**Branch 2**
QMS	Up or Down	-

---

[4] Reference to a 'branch' might seem odd given QMS. But we can think of this as a possibility where there is just the one branch. Furthermore, a referee correctly points out that nothing makes it branch 1 rather than branch 2 given QMS. I use the label for convenience. All that matters is that there is only one branch given QMS.

QME	Up	Down

Figure 2: Quantum mechanics

(This has the same structure as Uncertain Small Ball.) According to QMS, there is only one branch. One outcome will occur, which in this case is either Up or Down. The result of an observation will tell you which. According to QME, there are two branches. Making an observation tells you which branch you are on – the Up branch or the Down branch. Suppose you observe ‘Up’ (according to QME, the discovery is ‘I am in the Up branch’, where the ‘I’ refers to a post-measurement observer). Does the evidence confirm either QME or QMS? As QME is intended to be an interpretation of the same theory as QMS, the desired answer is: Neither is confirmed.[5] I will shortly offer an analyis that delivers this result.

But a tempting answer we will reject is that the evidence confirms QME. The argument for this confirmation of QME uses the non-self-locating evidence:

E_{-2} = There is a branch on which Up occurs

We then note that there is certain to be an Up branch given QME, but not given QMS; $P(E_{-2}|QME) = 1 > P(E_{-2}|QMS)$, so E_{-2} confirms QME. And notice that it doesn’t matter that Up was observed rather than Down. Both outcomes had a probability of 1 given QME, and less than 1 given QMS, so both must confirm MWI. We get the result that any possible evidence confirms QME. Call this result *Naïve Confirmation*.[6]

Naïve confirmation is absurd. But it is not surprising; as QME says that every outcome occurs, any outcome that isn’t entailed by QMS will confirm QME. And it gets worse. As branching is happening all the time, it would follow that we have overwhelming evidence in favour of QME. For example, suppose I see it is raining. Under any interpretation of quantum mechanics there are some quantum outcomes (however improbable) on which it wouldn’t have rained. According to QME, it was certain that rain would be observed (in some branch), but according to QMS it is less than certain. So such everyday observations are constantly confirming QME. On this reasoning, QME gets enormous confirmation without the need for

---

[5] See Saunders 1998, Vaidman 2002, Greaves 2004, Papineau 2004 and Wallace 2005. Particularly relevant is the debate, comparing the extent of the analogy between QME and Sleeping Beauty, in Peter Lewis 2007, 2009 and Papineau & Durà-Vilà 2009a, 2009b. My argument favours Lewis’s position.

[6] This name is adapted from Greaves’s Naïve *Conditionalization*. I have simplified the problem so it can be applied to a single probability model at one time.

modern physics. The Ancients could have worked out that they have overwhelming evidence for QME merely by realizing it was a logical possibility and observing the weather. Price (2006), worried by Naïve Confirmation, suggests that Bayesian confirmation theory cannot be applied to QME.

I think instead that all we need to do is take note of the self-locating nature of the evidence and the selection effects that come with it. The result wil be that the evidence favours neither QME or QMS.

We previously tried to express the evidence as:

E_{-2} = There is a branch on which Up occurs.

But this misses out the self-locating nature of the evidence. 'Up' has been found on *this* branch:

E_{-1} = Up occurs on this branch[7]

Now we need to add the selection effect with which to model the discovery of the self-locating belief. Was there a bias towards observing Up? That is, given that Up is instantiated, is it certain that I will observe Up? No. Remember the observation is made by a post-branching observer, for whom it was a possibility that Up was instantiated but not observed; this happens if QME is true and the other branch is Up. So the case should not be modelled with a biased procedure. Let's assume for now that the observer believes he is as likely to be in the Up branch as the Down branch given QME. (This assumption will be dropped shortly.) Then the case can be modelled with a random procedure. Knowledge of the selection effect could be expressed as part of the background knowledge, but for clarity we'll put it into the new evidence (and will do so for the other 3 cases):

E = I learnt that Up occurs in this branch by a random procedure.

---

[7] One might object that the only way to individuate 'this branch' is as 'the Up branch', and argue that this evidence has a probability of 1. But the probabilities are subjective, not objective. Even if this branch is necessarily the Up branch, this fact is not *a priori,* nor is it certain, so your prior credence in the evidence will be less than 1. 'This branch' is not functioning as a rigid designator in the relevant probability model, so I've tried to avoid using 'this'. I think the fact that self-locating terms like 'this' and 'I' are usually rigid designators is a source of confusion. See fn. 20 for a more intuitive example, Santorio (ms) and to some extent Bradley 2009.

This finally expresses the total evidence. It puts us in the top row of the table.

Now let's ask if it is certain that the Up result occurs at all? Is the property of being an Up branch certain to be instantiated? No. If QMS is true and Down occurs, then there is no Up branch. This puts us in the right column of the table.

	p is certain to be instantiated	**p might not be instantiated**
Random procedure	One Outcome confirmed Doomsday Argument	No confirmation **Quantum mechanics**
Biased procedure	No confirmation Sleeping Beauty	Two Outcomes confirmed Fine-Tuning

The procedure is random and the property discovered might not have been instantiated. More specifically, the probability that a randomly selected branch is an Up branch is the same given either QME or QMS.

$$P(E \mid QME) = P(E \mid QMS)$$

So we get the desired result that there is no confirmation of either QME or QMS when 'Up' is observed. (Similarly for observing 'Down'.)

There is a complication that must be dealt with – branches may have unequal *weights*. Weights are introduced to QME to allow for the possibility that – in the terminology of QMS - not all the outcomes are equally likely. The outcomes that are more likely according to QMS have greater weight according to QME. The upshot is that you are not always equally likely to be in the Up branch as the Down branch. If Up has a chance of 70% according to QMS, then the Up branch has a weight of 70% according to QME, which means a post-branching observer should assign a probability of 70% to being in the Up branch.

But this is no problem for our result that P(E | QME) = P(E | QMS). If P(E | QMS) = 70% due to the objective chances, P(E | QME) = 70% due to the weights. As long as the weights in QME equal the probabilities in QMS, E won't favour one over the other.[8]

---

[8] This removes the need for a Principle of Indifference defended in Bradley 2011.

Of course the problem remains for Everettians of justifying this use of weights. Why should we believe that we are on the Up branch with 70% certainty just because the Up branch has a weight of 70%? This is a central problem for Everettians that is beyond the scope of this paper (see the references in fn.5 for discussion). My claim is that if this match of subjective probabilities to weights can be justified, standard confirmation theory can be applied.

We'll now move on to an example where the procedure is biased, and see that confirmation of the Two Outcomes hypothesis results.

2.2 Fine-Tuning

If the fundamental constants of the universe had been much different from their actual values, life could not have existed. For example, if gravity had been a bit stronger, the universe would have collapsed in on itself moments after the big bang. If it had been a bit weaker, the universe would have flown apart so fast that molecules could never have been formed. The same holds for nearly all the other fundamental constants (see McMullin (1993)). (The initial conditions are also vital. For ease of exposition, I will take 'right constants' to include right initial conditions.) The existence of every living thing in the universe is balanced on a knife-edge[9]. Nevertheless, life exists.

Proponents of the fine-tuning argument claim that the existence of life requires an explanation[10]. One explanation is that there are many universes, and these universes have fundamental constants with different values. This is a hypothesis that has been independently suggested several times, from Wheeler's (1973) oscillating universes to Susskind's (2005) Landscape hypothesis. (Not all versions of these theories entail that the other universes have different constants, but I am interested in the versions in which they do. This restriction might reduce the prior probability of the hypothesis, but that doesn't affect my argument, which is about whether the hypothesis is confirmed by the evidence.)

We can model a simple version of the fine-tuning argument as having two hypotheses, with either one or two universes.

---

[9] I have found that some philosophers are hostile to this claim for reasons I find puzzling; physicists seem to take the claim as data. See Colyvan et al. (2005) for a technical argument that it is not improbable that the universe be fine-tuned, and Monton's (2005) response.

[10] These proponents include Craig (1988), Leslie (1989), Manson (1989), Swinburne (1990), van Inwagen (1993), Parfit (1998) and Monton (2005.).

	Universe 1	**Universe 2**
Universe (UV)	Right or wrong constants	-
Multiverse (MV)	Right or wrong constants	Right or wrong constants

Figure 3: Fine-tuning

(This has a similar structure to Uncertain Small Ball. A difference is that MV does not entail that the right constants are instantiated. But we'll see that MV does make the right constants more likely to be instantiated, and this is enough for MV to be confirmed.) Does our evidence confirm the Multiverse hypothesis? I will argue that it does.

To fit the Multiverse case into our framework, we must ask the same two questions. Firstly, what is the procedure by which our evidence been discovered? Well what exactly is the evidence? We might first try to express the evidence as:

E_{-2} = There is a universe with the right constants for life.

But this misses out the self-locating nature of the evidence. The right constants have been found in *this* universe:

E_{-1} = This universe has the right constants for life.

Now we need to add the selection effect that the self-location brings in.[11] Was there a bias towards our discovering a universe with the right constants for life rather than the wrong constants? That is, given the existence of a universe with the right constants, is it certain that we will observe a universe with the right constants? For now, we'll assume that it is (this assumption will be weakened shortly). So we can model this case as one in which there is a bias towards observing universes with the right constants.

---

[11] In Bradley 2009 I argue that Hacking 1987 and White 200 are right to include the self-locating evidence but get the selection effect wrong. See fn. 17.

> E = I have learnt that this universe has the right constants for life with a procedure biased towards universes with the right constants.

This bias puts us in the bottom row of the table.

Secondly, was the property of being a universe with the right constants for life certain to be instantiated? No. It was possible that no universe had the right constants for life.[12] This puts us in the right column of the table, so the two outcomes hypothesis (MV) is confirmed - $P(E|MV) > P(E|UV)$.

We'll look at a formal model in a moment, but it's worth going through an intuitive way to think about this argument.[13] Recall from Uncertain Small Ball that discovering, with a biased procedure, a property that is more likely to be instantiated given two balls than one, confirms the two ball hypothesis. Applied to the universe case, discovering, with a biased procedure, a property that is more likely to be instantiated given MV than UV, confirms MV. And we'll see that this sufficient condition for confirmation of MV holds on plausible assumptions.

Imagine first that God makes just one universe. Then the only chance of a universe with the right constants is if that one universe has the right constants. (Assume that each universe has a probability between 0 and 1, non-inclusive, of having the right constants.) Now suppose God makes a second universe. There is now a second chance for the right constants to be instantiated. If the probability of this second universe existing and having the right constants is independent of the first having the right constants, then the overall probability of the right constants being instantiated must rise. As the biased procedure ensures that a universe with the right constants is observed whenever it is instantiated, this means that the overall probability of observing the right constants must rise. Thus, discovering the right constants confirms the Multiverse hypothesis.[14]

	p is certain to be instantiated	**p might not be instantiated**

---

[12] Sober 2003 disputes this. See Monton 2005, Weisberg 2005 and Kotzen (ms) for responses.

[13] I am grateful to Mike Titelbaum for an exchange on what follows in this section.

[14] Related discussions can be found in Bostrom 2002, Juhl 2005 and Oppy 2006.

Random procedure	One Outcome confirmed Doomsday Argument	No confirmation Quantum mechanics
Biased procedure	No confirmation Sleeping Beauty	Two Outcomes confirmed **Fine-Tuning**

We can now weaken the assumption that the procedure is biased.[15] We can run through the argument of the previous paragraph but weaken the last step regarding the procedure. Grant that the existence of a second universe increases the overall probability that the right constants are instantiated. It's plausible that an increase in the overall probability that the right constants are instantiated also increases the overall probability that we will exist[16] and therefore observe the right constants. If so, we are more likely to observe the right constants given MV than UV. I'll now give a more formal argument.

b = The probability that any given universe has the right constants for life.
L = At least one universe has the right constants for life.
UV = There is exactly one universe.
$P(L \mid UV) = b$
MV = There are exactly two universes.
C = Universe 1
D = Universe 2
CL = Universe 1 has the right constants for life
DL = Universe 2 has the right constants for life

Assume that each possible universe has the same chance of having the right constants, and that these probabilities are independent of each other:

$$P(CL) = P(DL) = P(CL|DL) = P(DL|CL) = b$$

So,

$$P(L \mid MV) = 1-(1-b)^2$$

---

[15] Indeed we must - there might have been universes with the right constants, but none that contain us. For example, the exact planetary conditions required for life to begin might not have occured.

[16] This is the step that fails if the procedure is random. With a random procedure, what matters isn't the overall probability that the right constants are instantiated, but the probability that a given universe has the right constants. See fn. 16.

$$= b+(b-b^2)$$
$$= b+b(1-b)$$

Assuming only that b is between 0 and 1, it follows that b and (1-b) will both be positive, so b(1-b) will be positive, so b+b(1-b) > b, so P(L | MV) > P(L | UV). So L supports MV.

But L may be true, yet we don't observe the right constants. That is, the procedure is not biased. To see how the argument will still go through, first assume that the procedure is biased, so that if some universe has the right constants, we observe the right constants.[17] So P(L) = P(We learn L)[18], and we can substitute P(We learn L) for P(L).

$$\begin{aligned} P(\text{We learn } L \mid UV) &= b \\ P(\text{We learn } L \mid MV) &= 1-(1-b)^2 \\ &= b+(b-b^2) \\ &= b+b(1-b) \end{aligned}$$

As above, b+b(1-b) = P(We learn L | MV) > P(We learn L | UV) = b, so We Learn L supports MV.

Now let's drop the assumption that the procedure is biased. We can do this by adding another parameter connecting a universe having the right constants for life and us observing it:

> s = The probability that we observe the right constants given that there is exactly one universe with the right constants.

Assuming that we are as likely to exist in either possible universe the effect this extra parameter has given UV is straightforward:

---

[17] White 2000, following Hacking 1987, denies this on the grounds that we could not have existed in any universe other than the one we are actually in, say, C. This has the same effect as a random procedure would - there would be no confirmation. In Bradley 2009 I argue that we could have existed in other universes, and even if we couldn't have, the Multiverse argument would still go through; granting some independence assumptions, the more universes there are, the greater the chance that C exists, so the greater the chance that we will exist. This section makes explicit some assumptions implicit in that paper.

[18] This also requires that if we observe the right constants, some universe has the right constants. This is true assuming either veridical observations or, more weakly, no 'freak observers' in universes with the wrong constants (Bostrom 2002 ch.5)

P(We learn L | UV) =sb.

How does this compare to MV?

Assume that if one universe has the right constants then the existence of another with the wrong constants doesn't change the probability that we observe the first:

P(We learn L | CL.~DL) = P(We learn L | ~CL.DL) = s

If instead no universe has the right constants, the probability we observe a universe with the right constants is 0:

P(We learn L |MV. ~CL.~DL) = 0

What about if both exist and have the right constants? We cannot exist in both universes. To make the case as difficult as possible for the multiverse theorist, let's say that the probability of us observing a universe with the right constants given two universes with the right constants is also s. So the probability of us observing a universe with the right constants is no greater given two universes with the right constants than given one universe with the right constants:

P(We learn L | CL.DL) = s

These four possibilties for MV have the following weights:

P(CL.DL) = b^2
P(CL.~DL) = b(1-b)
P(~CL.DL) = (1-b)b
P(~CL.~DL) = $(1-b)^2$

So P(We learn L | MV) = P(We learn L | CL.DL).P(CL.DL) +
P(We learn L |CL.~DL).P(CL.~DL) +
P(We learn L | ~CL.DL). P(~CL.DL) +
P(We learn L | ~CL.~DL). P(~CL.~DL)
= sb^2 + sb(1-b) + sb(1-b) + 0
= sb + (sb - sb^2)
= sb + sb(1-b)

Assuming only that sb and b are between 0 and 1, it follows that sb and (1-b) will both be positive, so sb(1-b) will be positive, so sb + sb(1-b) > sb, so P(We learn L | MV) > P(We learn L

| UV). So ‘We Learn L' confirms MV. We can conclude that the discovery of life does confirm the Multiverse hypothesis.

We’ll now move on to cases where the property discovered is certain to be instantiated, modelled by Certain Small Ball and the left side of the table.

2. 3 The Doomsday Argument

The Doomsday Argument claims that from the discovery of your birth rank, and the assumption that you are an average observer (in a sense to be explained), the hypothesis that there will be relatively few observers is confirmed. That is, you get evidence that Doomsday will occur sooner rather than later[19]. We can generate a simple Doomsday-style argument by imagining that there is either a total of one or a total of two people in the universe. These hypotheses are mutually exclusive and exhaustive.

H1: There is one person in the universe.
H2: There are two people in the universe.

Each person created is put in an isolation cubicle, so they do not know if anyone else exists. They are numbered: person 1, and if he exists, person 2. Suppose you find yourself existing in this scenario (which you are told about). There are three possible states you might be in.

	Self-Locating Possibility 1	**Self-Locating Possibility 2**
H1	Person 1	
H2	Person 1	Person 2

Figure 4: The Doomsday Argument

---

[19] Carter 1974, Leslie 1989, 1996

(This has the same structure as Certain Small Ball.) Suppose you learn that you are person 1. Does this confirm either H1 or H2? I will argue that it confirms H1, as the Doomsday Argument claims. We must ask the same two questions.

Firstly, by what procedure have I discovered the evidence? Well what exactly is the evidence? We might first try to express the evidence as:

E_{-2} = Person 1 exists.

But this misses out the self-locating nature of the evidence:

E_{-1} = I am person 1

Now we need to add the selection effect the self-location brings in. Was there a bias towards discovering I am person 1 rather than person 2? That is, given the existence of person 1, is it certain that I learn that I'm person 1? No; for all I knew, I could have been person 2[20]. Furthermore, I've no reason to think I'm person 1 rather than person 2. In the Doomsday Argument literature, this is often expressed as the assumption that we should expect to be average[21]. Let's assume for the moment that given H2, you are as likely to be person 1 as person 2. The case can then be modelled by a random procedure. (We'll see in a moment that this randomness assumption can be greatly weakened.)

E = I have learnt that I am person 1 by a random procedure.

The random procedure puts us in the top row of the table (below).

Secondly, is the property discovered certain to be instantiated? Yes. The property of being person 1 is certain to be instantiated because person 1 exists given either H1 or H2. This puts us in the left column of table 2. Obviously H1 entails E; and H2, due to the random procedure, assigns a probability of 1/2 to E:

---

[20] One might object that the only way to individuate the observer is as person 1, and argue that this has a chance of 1. But the probabilities are subjective, not objective. Even if you are necessarily person 1, this is not *a priori,* nor is it certain, so your prior credence will be less than 1. So 'I' is not functioning as a rigid designator in the relevant probability model. See fn.7.

[21] Dieks 1992, Eckhardt 1993. The assumption we make in the next sentence is defended by Elga 2004.

$$P(E|H2) = 1/2 < P(E|H1) = 1$$

We can conclude that the One Outcome hypothesis, H1, is confirmed, thus validating the Doomsday Argument.

We can now weaken the assumption that the procedure is random. H1 is confirmed whenever $P(E|H2) < P(E|H1) = 1$. This condition holds whenever the probability of being person 1 is less than 1, which holds whenever there is a non-zero probability that you will be person 2. So the only assumption we need for the Doomsday Argument is that there is non-zero probability that you are person 2.[22] We'll now move onto a case where the procedure is biased, and find that neither hypothesis is confirmed.

	p is certain to be instantiated	p might not be instantiated
Random procedure	One Outcome confirmed **Doomsday Argument**	No confirmation Quantum mechanics
Persistent procedure	No confirmation Sleeping Beauty	Two Outcomes confirmed Fine-tuning

2.4 Sleeping Beauty

In the Sleeping Beauty problem (Elga 2000), the self-locating variable is the time.

> 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. 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.

---

[22] See Bradley and Fitelson (2003)

	Monday	**Tuesday**
Heads	Awake	Asleep
Tails	Awake	Awake

Figure 5: Sleeping Beauty

(This has a similar structure to Certain Small Ball, but notice that Heads and Tails both have the same population here (two). I'll explain why this won't affect our analysis at the end of this section.) When Beauty is woken, is Tails confirmed?

Some say no; her credence in Heads should stay at 1/2. Call these *Halfers.*

Some say yes; her credence in Heads should fall to 1/3. Call these *Thirders.*

Some thirders think that Beauty learns new evidence on being woken, and updating on this new evidence confirms Tails (Weintraub 2004, Horgan 2007). But I will argue that this position is mistaken. To determine whether the evidence confirms Tails, we must ask the same two questions.

Firstly, by what procedure has Sleeping Beauty discovered the evidence? Well what exactly is her evidence? We might first try to express the evidence as:

E_{-2} = There is a day on which I am woken.

But this misses out the self-locating nature of the evidence:

E_{-1} = I am woken today.

Now we need to add the selection effect that the self-location brings in. Was there a bias towards discovering a waking day rather than a sleeping day? That is, given the existence of a day on which I'm awake, is it certain that I observe a day on which I'm awake? Yes.[23] If there is a day on which I'm woken then I must observe a day on which I'm woken. So the case can be modelled using a procedure that is biased towards waking days.

E = I have learnt that I am woken today by a procedure biased towards waking days.

The biased procedure puts us in the bottom row.

Secondly, is the property of being a day Beauty is woken certain to be instantiated? Yes. Given either Heads or Tails, there is a day Beauty is woken. This puts us in the left column. We have a property that was certain to be instantiated being discovered by a biased procedure, so neither Heads nor Tails is confirmed.

P(E | Heads) = P(E | Tails) = 1.

	p is certain to be instantiated	p might not be instantiated
Random procedure	One Outcome confirmed Doomsday Argument	No confirmation Quantum mechanics
Biased procedure	No confirmation **Sleeping Beauty**	Two Outcomes confirmed Fine-tuning

This answer coheres with the halfer position. There are many and varied arguments for being a thirder that I am not dealing with of course.[24] My argument here is that 'I'm awake' does not confirm Tails in any straightforward way.

Why didn't it matter that the population was the same size given either Heads or Tails? Because if we have evidence that is certain to be instantiated (given either hypothesis), and certain to be found (given the biased procedure) then the probability of observing the evidence must be 1. Such evidence can't confirm either hypothesis, no matter which other objects there may be in the population.

---

[23] Note that P(E | Tails) = 1 just given the set-up of the case (see figure 5). The biased procedure ensures that P(E | Heads) =1.

[24] See Elga 2000, Dorr 2002, Hitchcock 2004, Draper and Pust 2008 and Titelbaum 2008 for a selection.

Part 3: Alternative Theories

This analysis reveals the core common structure of all four problems. It follows that a theory of self-locating belief based on any one of these problems can and should be applied to the other cases.[25] I will now briefly describe three theories regarding self-locating beliefs, each introduced in response to one of the problems, which give problematic results when applied to another.

3.1 The Self-Indication Assumption

In response to the Doomsday Argument, some writers (Kopf, Krtous and Page 1994, Bartha and Hitchcock 1999, Olum 2002) have endorsed the Self-Indication Assumption:

> (SIA) Given the fact that you exist, you should (other things equal) favour hypotheses according to which many observers exist over hypotheses on which few observers exist. (Bostrom 2002)

One motivation for this principle is that it perfectly counter-balances the Doomsday Argument. That is, the shift towards fewer observers generated by the Doomsday Argument is perfectly cancelled out by the confirmation of more observers by SIA.

The problem is that SIA is not just applicable when considering how many observers will exist in the future on this planet. It also seems to apply when considering whether the Everettian Interpretation (QME) or the Stochastic Interpretation (QMS) is correct. There will be many orders of magnitude more observers given QME than QMS. So SIA implies that you should strongly favour QME over QMS, just on the basis of your existence.

Defenders of SIA may try to reply that there is an equivalent of the Doomsday Argument that confirms QMS and is cancelled out by the confirmation of QME. But there isn't. Indeed, I argued earlier that neither QME nor QMS is confirmed by discovering which branch you are on, so there is no shift to be cancelled out. This contrasts with the Doomsday

---

[25] Dieks 2007 and Peter Lewis 2007, 2010, make similar connections and come to a different conclusion. The following section on the Self-Indication Assumption could be seen as a response.

Argument, where discovering your birth rank confirms that there are fewer observers.[26] We are left with a huge shift in favour of QME based solely on our existence.[27]

3.2 Titelbaum

Mike Titelbaum (2008, forthcoming) develops a theory of confirmation regarding self-locating belief in the context of Sleeping Beauty and runs into similar issues. He argues that the problem for Sleeping Beauty is that she has no non-indexical way to refer to 'today'. He proposes fixing the problem by giving the days some uniquely distinguishing feature. He does this by modifying the setup so that on each day Beauty is awake she sees a coloured paper. If she is woken on both days, she sees a red paper on one day and a blue paper on the other (there is never a correlation between the day and the colour of the paper). If she is woken on only one day, there is a 50% chance she sees a red paper and a 50% chance she sees a blue paper. Beauty, as always, knows all this. Now when she wakes up and sees, say, a red paper, she can update on the non-indexical 'a red paper is observed'. As it is certain a red paper will be observed given two wakenings but only has a 50% given one awakening, the two wakenings hypothesis (Tails) is confirmed. Thus we have an argument for the Thirder position.

But the same strategy applied to QME gives us an implausible result. We already have uniquely identifying descriptions of the branches – one is Up and the other is Down. Applying Titelbaum's theory, we can update on 'an Up branch is observed'. 'An Up branch is observed' is certain given QME but is less than certain given QMS, so QME is confirmed by Up. The same applies to Down of course. So QME is confirmed whatever the outcome of the experiment happens to be. Titelbaum is effectively committed to the same bad result as Naïve Confirmation. And we saw earlier that as branching is happening all the time, the Ancients could have worked out that they have overwhelming evidence for QME merely by realizing it was a logical possibility and observing the weather. Although I have focussed on Titelbaum, many arguments for being a Thirder in Sleeping Beauty transfer over into Naïve Confirmation in QME.[28]

3.3 Meacham

---

[26] See Bradley 2005 for a discussion of the special feature of birth rank.

[27] This is an actualized version of Bostrom's 2002 Presumptuous Philosopher thought experiment.

[28] See Peter Lewis 2007.

Chris Meacham (2008)[29] argues that learning a self-locating belief should never alter credence in a non-self-locating belief (as long as the non-self-locating belief isn't eliminated). He also develops this theory in the context of the Sleeping Beauty problem. He argues that Beauty's credence in Heads should remain 1/2 when she is awoken, and should remain at 1/2 even if she were to learn that it is Monday. Meacham uses the case to defend general norms of belief change for self-locating belief.

But the theory looks problematic when we extend it to the other cases. For example, when we apply these norms to the Doomsday Argument, we get the result that learning that you are person 1 shouldn't have any impact on the probability of there being 1 person in total or 2 people in total. After all, you have learnt something purely self-locating – it was certain that someone would be person 1, you've just learnt that *you* are person 1.

As a referee points out, it may be too strong to say that this result is problematic – the Doomsday Argument is after all an argument that most want to reject, and perahps Meacham has shown us how to reject it. But my intuition at least is that Meacham's suggestion is far less plausible when applied to the Doomsdsay Argumen that when applied to Sleeping Beauty. And I think it is telling that none of the numerous critics of the Doomsday Argument ever suggested anything resembling Meacham's approach. This suggests to me that Meacham's solution is *ad hoc* and does not extend naturally to related cases.

Even if you think that Meacham's approach does extend to the Doomsday Argument, then perhaps you accept my larger point, which is one about strategy. Given the close connection between the four cases, we should not examine them in isolation. A theory based on one case should be tested against the others. There may be significant disanalogies, and this will come out in the analysis. But the similarities between the four cases give us more weapons in our armoury when thinking about self-locating belief, and raises problems for several theories suggested so far in the literature.

Conclusion

The four problems discussed are all cases where an agent learns where she is located. I have argued that there is nothing mysterious about this kind of learning. I have applied the traditional view of confirmation, where E confirms H iff P(E|H) > P(E|-H) and defended the resulting positions on each of the four problems. I have argued that all four problems have the same

---

[29] Meacham (2010) has since modified his position. His earlier view suffices for the point I want to make.

structure. The differences come from the selection procedure and whether the property discovered was certain to be instantiated. Using very simple models of the cases and no controversial assumptions about confirmation or self-locating evidence, I have defended four positions – the Doomsday Argument, the halfer position in Sleeping Beauty, the Fine-tuning argument, and the applicability of confirmation theory to the Everett interpretation of quantum mechanics. Someone who wanted to endorse one of these arguments and not another would have to demonstrate the disanalogy between the cases. More generally, the structural similarity between the cases shows that our theories should be plausible when applied to all four of these cases, and this will restrict our options in a hopefully productive way.[30]

References

Bartha and Hitchcock (1999) 'No One Knows the Date of the Hour: An Unorthodox Application of Rev. Bayes's Theorem' *Philosophy of Science*, 66 (Proceedings): p. S229-S353.

Bradley, D. & Fitelson, B. (2003) Monty Hall, Doomsday and Confirmation, *Analysis* 63, 23-31

Bradley, D. (2005) No Doomsday Argument Without Knowledge of Birth Rank *Synthese* 144, 91-100

-- (2009) Multiple Universes and Observation Selection Effects *American Philosophical Quarterly* 46, 61-72

-- (2010) 'Conditionalization and Belief De Se' *Dialectica* Vol. 64, N° 2, pp. 247–250

(2011) 'Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty' *British Journal for Philosophy of Science* 62 (2):323-342.

-- (forthcoming a) Self-Location is No Problem for Conditionalization *Synthese*

-- (forthcoming b) 'Dynamic Beliefs and the Passage of Time' in Feit and Capone (ed.s) *Attitudes 'de se': linguistics, epistemology, metaphysics*

Bostrom, N. (2002) Anthropic Bias: Observation Selection Effects in Science and Philosophy. New York : Routledge.

---

[30] I have had helpful exchanges with dozens of people on the material in this paper over the years. For discussion of recent predecessors of this paper I am grateful to Mike Titelbaum, audiences at the universities of Manchester, York, St. Andrews, British Columbia, Sydney, Delaware and CCNY, and two refereees for this journal.

Carter, B. (1974) ‘Large Number Coincidences and the Anthropic Principle in cosmology’ *Confrontation of Cosmological Theories with Data* M. S. Longair. Dordrecht, Reidel: 291-8.

Colyvan, M., Garfield, J. L., Priest, G. (2005) 'Problems with the Argument from Fine Tuning', *Synthese*, Vol. 145, No. 3 p. 325–38.

Craig, W. L. (1988) ‘Barrow and Tipler on the Anthropic Principle vs. Design’ *British Journal for the Philosophy of Science* 38: 389-395.

Dieks, D. (1992) ‘Doomsday - Or: the Dangers of Statistics’ *Philosophical Quarterly*, 42 (166): 78-84.

-- (2007) ‘Reasoning about the future: Doom and Beauty’ *Synthese* 156:427–439

Dorr, C. (ms) ‘A Challenge for Halfers’

Eddington, A. (1939) The Philosophy of Physical Science. Cambridge: Cambridge University Press.

Elga, A. (2000) ‘Self-Locating Belief and the Sleeping Beauty Problem’, Analysis. 60:143-7

– (2004) ‘Defeating Dr. Evil with Self-Locating Belief’ *Philosophy and Phenomenological Research* 69(2).

Greaves, H. (2004) ‘Understanding Deutsch’s Probability in a Deterministic Universe’. *Studies in History and Philosophy of Modern Physics*. 35: 423–456

-- (2007) ‘Probability in the Everett Interpretation: A Solution to the Epistemic Problem’ *Studies in History and Philosophy of Modern Physics* 38(1), March pp.120-152.

Groisman, B. (2008) ‘The End of Sleeping Beauty’s Nightmare’ *British Journal for the Philosophy of Science* 59: 409-416.

Hájek, A. (2007) ‘The Reference Class Problem is Your Problem Too’, *Synthese* 156: 185-215. 2007.

Halpern, J. (2004) ‘Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems’, Proceedings of the Twentieth Conference on Uncertainty in AI, 226-34

Hitchcock, C. (2004) ‘Beauty and the Bets’ *Synthese* 139(3): 405–20.

Horgan, T. (2004) ‘Sleeping Beauty Awakened: New Odds at the Dawn of the New Day’ *Analysis* 64: 10-21

– (2007) ‘Synchronic Bayesian updating and the generalized Sleeping Beauty problem’ *Analysis* 67:293, 50–59

Horwich, P. (1982) *Probability and Evidence,* Cambridge: Cambridge University Press.

Hutchison, K. (1999) ‘What Are Conditional Probabilities Conditional On?’ *British Journal for Philosophy of Science* 50, 665-95.

Juhl, C. (2005) ‘Fine‐tuning, Many Worlds, and the 'Inverse Gambler's Fallacy'’ *Nous* 39 (2): 337–347.

Kotzen, M. (forthcoming) ‘Multiple Studies and Evidential Defeat’ *Nous*

-- (ms) ‘Selection Biases in Likelihood Arguments’

Kopf, T., P. Krtous, and D.N. Page (1994) ‘Too soon for doom gloom’. Physics preprint, 1994. gr-gc/9407002(v3, 4 Jul.).

Leslie, J. (1989) ‘Universes’, London: Routledge.

– (1996) The End of the World: The Science and Ethics of Human Extinction. Routledge.

Lewis, P. J. (2007) ‘Quantum Sleeping Beauty’ *Analysis* 67(293): 59–65

-- (2009) ‘Reply to Papineau and Durà-Vilà’ *Analysis* 69: 86-89

-- (2010) ‘A Note on the Domosday Argument’ *Analysis* 70(1) p.??

Manson and Thrush (2003) ‘Fine-Tuning, Multiple Universes, and the "This Universe" Objection’ *Pacific Philosophical Quarterly* 84 (1): 67–83.

Manson, N. A. (1989) *Why Cosmic Fine-Tuning Still Needs to be Explained*. PhD dissertation, Philosophy, Syracuse University.

McMullin, E. (1993) ‘Indifference Principle and Anthropic Principle in Cosmology’, *Studies in the History and Philosophy of Science*, 24: 359–89.

Meacham, C. (2008) ‘Sleeping Beauty and the Dynamics of *De se* Belief’ *Philosophical Studies* 138:2 p. 245-269.

Monton, B. (2005) ‘God, Fine-Tuning and the Problem of Old Evidence’ *British Journal for Philosophy of Science*

Olum, K. (2002) ‘The Doomsday Argument and the Number of Possible Observers’, *Philosophical Quarterly* 52 (207): 164-84.

Oppy, G. (2006) *Arguing About Gods.* Cambridge University Press.

Papineau, D. (2004) ‘David Lewis and Schrodinger’s Cat’ *Australasian Journal of Philosophy* 82: 153-169

Papineau, D. and V. Durà-Vilà. (2009a). ‘A thirder and an Everettian: a reply to Lewis’ *Analysis* 69: 78-86.

-- (2009b) ‘Reply to Lewis: metaphysics versus epistemology’ *Analysis*, 69: 89 – 91.

Parfit, D. (1998) *Why anything? Why this?* London Review of Books, 1998. Jan 22, 24-27.

Price, Huw (2006) ‘Probability in the Everett World: Comments on Wallace and Greaves’. Available online from http://arxiv.org/abs/quant-ph/0604191

Salmon, W. C. (1975) ‘Confirmation and relevance’ in G. Maxwell and R. M. Anderson, Jr. (eds.) *Minnesota Studies in the Philosophy of Science* (6) p.3-36

Santorio, P, (ms) ‘Reference and Monstrosity’

Saunders, S. (1998) ‘Time, Quantum Mechanics and Probability’. *Synthese*, 114: 373-404.

Schwarz, W. ms ‘Changing Minds in a Changing World’

Sober, E. (2003) ‘The Design Argument’ in *God and Design: The Teleological arguments and Modern Science* (ed.) Manson, N., London: Routledge

Stuart, A. (1962) *Basic Ideas of Scientific Sampling,* London: Charles Griffin.

Susskind, L. (2005) *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design,* Little: Brown & Co.

Swinburne, R. (1990) ‘Arguments from the Fine-Tuning of the Universe’, in J. Leslie (ed.), *Physical Cosmology and Philosophy*, New York: MacMillan, 160–79.

Titelbaum, M. (forthcoming) ‘The Relevance of Self-Locating Beliefs’ *Philosophical Review*

Vaidman, L. (1998), “On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory”, *International Studies in the Philosophy of Science* 12: 245–261.

- (2002), “Many-worlds interpretation of quantum mechanics”, in Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online at http://plato.stanford.edu/archives/sum2002/entries/qm-manyworlds/.

van Inwagen, P. (1993) *Metaphysics*, Boulder, CO: Westview Press.

Vos Savant, M. (1997) *The Power of Logical Thinking*, New York: St. Martin’s Griffin

Wallace, D. (2006) ‘Epistemology Quantized: circumstances in which we should come to believe in the Everett interpretation’. Preprint available online from http://philsci-archive.pitt.edu/archive/00002839/

Weintraub, R. (2004) ‘Sleeping Beauty: A Simple Solution’ *Analysis* 64: 8-10

Weisberg, J. (2005) ‘Firing Squads and Fine-Tuning: Sober on the Design Argument’ *British Journal for the Philosophy of Science* 56: 4

Wheeler, J. A. (1977) in *Foundational Problems in the Special Sciences*, Reidel, Dordrecht p.3-33.

White, R. (2000) ‘Fine-tuning and Multiple Universes’ *Nous* 34 (2): 260-276