Belief Dynamics across Fission

Wolfgang Schwarz
Draft, 17/11/2011

Abstract. When an agent undergoes fission, how should the beliefs of the fission products be related to the pre-fission beliefs? This question has recently drawn some attention in the context of Everettian quantum mechanics, but it is of independent philosophical interest. Among other things, it provides counterexamples to popular principles concerning self-locating indifference, peer disagreement, and the relationship between objective chance and rational credence. It also supports the Lewisian "halfer" solution to the Sleeping Beauty problem, and illustrates the important difference between evidential probability and rational belief. Finally, this paper completes the general model of belief update presented in [Schwarz 2011], which did not apply to cases of fission.

1 The problem

Fred's home planet, *Sunday*, is surrounded by two moons, *Monday* and *Tuesday*. Tonight, while Fred is asleep, his body will be scanned and destroyed; then a signal is sent to both Monday and Tuesday where he will be recreated from local matter.

A lot of ink has been spent on how to describe scenarios like this. Should we say that Fred will find himself both on Monday and on Tuesday? Which of the persons awakening on the two moons is identical to the person going to sleep on Sunday? In this paper, I want to look at a different question: what should Fred's successors believe when they awaken on Monday and on Tuesday? More precisely, how should their beliefs be related to Fred's beliefs before he went to sleep on Sunday?

The two issues are independent. For the present topic, it does not matter whether the two "successors" are identical to Fred, either temporally or absolutely. Suppose, for example, that Fred does not survive the double teleportation, so that two new persons will come into existence on Monday and on Tuesday. We can still ask how the beliefs of these persons should be related to the beliefs of Fred. Imagine you are designing a population of intelligent amoebae that regularly undergo fission. What update process would you implement for the amoebae's beliefs so as to make optimal use of the information they have collected?

The story of Fred gets more interesting if he doesn't know what is going to happen. Suppose Fred learns that a fair coin will be tossed and that the signal to Tuesday is cut

if the coin lands heads. *We* know that the coin lands tails and the signal isn't cut, but Fred doesn't. Now how confident should his successors be that they are on Monday? What should they believe about the outcome of the coin toss?

The scenario bears an obvious resemblance to the Sleeping Beauty problem. One initially plausible answer is that the successors should assign credence 1/2 to heads and 3/4 to Monday. Another is that their credence in heads should be 1/3 and their credence in Monday 2/3. We will see, however, that there are also some disanalogies between the two cases. I will argue that "halfing" is the right answer in the case of Fred, whereas the Sleeping Beauty problem allows for different interpretations, some of which support halfing and some thirding.

This is one reason to care about belief update across double teleportation and other cases of fission: the topic can shed new light on the Sleeping Beauty problem. But there are further reasons. For one thing, Fred's predicament is less far-fetched than is may at first appear. According to the Everett interpretation of quantum mechanics (in its most popular form), what are commonly regarded as chance events are really *branching* events in which every possible outcome determinately occurs on some "branch" of the universe. Agents who witness such an event partake in the branching, each of their successors witnessing one particular outcome. Now suppose you give positive credence x to the Everett hypothesis. If that hypothesis is true, you are frequently subject to fission. Given your divided state of mind, how should your beliefs evolve?

There are also more theoretical reasons to care about belief dynamics across episodes of fission. I will argue that fission scenarios provide counterexamples to popular principles concerning self-locating indifference, peer disagreement, and the relationship between objective chance and rational credence. They also put pressure on several general models of belief update, including the one I presented in [Schwarz 2011]. I will begin by reviewing this account.

2 Conditioning and self-location

When Fred's Monday successor wonders whether he is on Monday or on Tuesday, what he lacks is not objective information about the universe, but "self-locating" information about himself. Following [Lewis 1979], it has become customary in the Bayesian tradition to model self-locating information in terms of *centred possible worlds*. A centred possible world (for short, a *world*) is a maximally specific way things could be for a subject at a time. A set of centred worlds is a *(centred) proposition*.

A useful heuristic is to think of centred propositions as properties. Fred's Monday successor, for example, will give some degree of belief to *being on Monday* and some to *being on Tuesday*. A world is a maximally specific property. Worlds also contain information about the universe as a whole. Suppose Fred is uncertain whether space-time

is Euclidean. Then he gives some degree of belief to maximally specific properties that entail *being such that space-time is Euclidean*. If two worlds entail the same information about the universe as a whole, I will say that they are *worldmates*. A maximal proposition all members of which are worldmates of one another is an *uncentred world*.[1]

Ordinary objects trace a path through the space of centred worlds. Consider Napoleon. Initially, in 1769, he had properties like *living on Corsica* and *being called Napoleone*; later he lost these properties and acquired new ones, until he eventually had such properties as *being 51 years old* and *living on St. Helena*. At each point in his life, the totality of Napoleon's properties constituted a centred world. The complete history of Napoleon is therefore a sequence of possible worlds: a trajectory through logical space.[2]

To keep distracting technicalities at bay, I will pretend for most of this paper that there are only finitely many worlds. This means that every intermediate point in Napoleon's trajectory has a determinate successor, unlike the points on the real number line. It also means that we can represent an agent's beliefs by a simple probability distribution over the entire space of worlds, without having to worry about infinite additivity and non-measurable sets.

The Bayesian rule of *conditioning* can be interpreted as a rule for how an agent's probabilities should evolve over time. Let w_1, w_2 be subsequent positions in an agent's trajectory. Conditioning says that the probabilities at w_2 equal the probabilities at w_1 conditional on the total evidence E_2 received at w_2:

$$P_2(A) = P_1(A/E_2).$$

In the present context, however, this is not a sensible rule. It ignores the fact that agents constantly change their position in logical space. Suppose at w_1 you believe that your present position has a certain property A. Later, at w_2, you learn that your new position has property E. Should this make you believe that your *new* position has A to the extent that you previously believed that your *old* position has A conditional on it having E? Clearly not, unless you have reason to believe that A and E don't change their truth-value between the two positions. If you know that Smith is either in his office or in the pub, and then you learn that Brown is not in the pub, you can only conclude that Smith is in his office if you have reason to believe that Smith and Brown are together.

1 Many authors take the notion of an uncentred world as primitive and define centred worlds as triples of an uncentred world, an individual and a time. I think it is more natural to take centred worlds as primitive. This also avoids the question whether world-time-individual triples contain enough information to determine e.g. directions, and it does not run into the obvious problems for the triples account if the relevant individual is a time-traveler or a multi-headed dragon.

2 There is a lively debate in metaphysics about whether an individual x having a property F at a time t should be analysed in terms of a tensed instantiation relation between F, t and x, or in terms of an untensed relation between F and x's time-slice at t. For our purposes, we don't need to take sides. Either way, Napoleon will have had properties like *living on Corsica* early in his life and *living on St. Helena* later.

As a general rule for updating beliefs, conditioning needs to be revised. The revised rule should still determine the agent's new beliefs based on the previous beliefs together with the new evidence. But it must take into account the fact that propositions can change their truth-value between w_1 and w_2.

Here it helps to keep in mind two facts. First, we can assume that w_2 is an immediate successor of w_1. If we want to determine an agent's credence at w_2 based on their credence at w_1 and the evidence at w_2, we cannot in general allow there to be further points in between w_1 and w_2 – at least no points at which relevant evidence is received. For if the agent receives evidence in between w_1 and w_2, then their credence at w_2 should be sensitive to this evidence, and we can't assume that all such evidence is part of the evidence at w_2, if only because what is learnt at the intermediate point may already be false at w_2.

The second fact to keep in mind is that worlds are maximally specific. A world does not only contain information about the present, but also about the past and the future. In particular, if a world w lies on the trajectory of some agent, then it fully determines which other worlds, in which order, lie on the trajectory of this agent. Consider Napoleon's position in logical space on New Year's eve 1805. This position settles not only what Napoleon is doing right at that time, but also what everybody else is doing at every other time. Moreover, it entails that Napoleon himself will die on St. Helena in 1824, and that the world centred on this event is a distant descendant of the present world.

Given these two facts, it is not hard to find the required amendment to conditioning. We simply need to add an operation to the update process that *shifts* the probability of all previously possible worlds to their successors. This shifted probability is then conditioned on the new evidence. If the agent is omniscient, so that all their credence at w_1 is concentrated on the single world w_1, then the update moves the credence to the successor of w_1, which is w_2. If the agent's credence is divided 2:1 between two worlds w and v, and the new evidence is compatible with either successor, then the new credence should be divided 2:1 between the successor of w and the successor of v. In general, an update consists in moving the probability of each world to its successor and then ruling out the worlds incompatible with the new evidence. This two-stage process has long been used in computer science.[3] In philosophy, it has only recently been rediscovered by Christopher Meacham [2010b] and myself [2011].

Let me spell out the amended rule in more detail. Suppose, for now, that every world with positive probability at w_1 has exactly one successor. That is, every such world lies on a trajectory where it is succeeded by a unique other world. Define the shifting operator '$\succ$' (read 'next') so that $\succ w$ is true at a world v iff w is a successor of v. More generally, for any proposition A, let $\succ A$ be true at w iff A is true at a successor of w.

3 See [LaValle 2006], part III, for a textbook presentation in the context of stochastic control theory, and [Boutilier 1998] for an application of the same ideas to the framework of [Alchourrón et al. 1985].

Given a probability function P, define $P^{\succ}$ so that $P^{\succ}(w) = P(\succ w)$ for every world w. This is the *shifted* probability function under which the probability of each world has been moved to its successor. Finally, the new probability is the shifted previous probability conditional on the new evidence:

$$P_2(A) = P_1{}^{\succ}(A/E_2).$$

Call this amended form of conditioning *shifted conditioning*. Instead of first shifting and then conditioning on E, we could also first condition on $\succ E$ and then shift, like so:

$$P_2(A) = P_1(\succ A/\succ E).$$

The result is the same.

Let me say a bit more on the idea of a "next" world. Formally, the propositional operator '$\succ$' works much like 'in five minutes'. For example, if all the probability goes to worlds whose successors are located on Monday, then $P(\succ \mathit{Monday}) = 1$. One reason for not using an objective time interval in place of '$\succ$' is that the distance between a world and its successor can vary widely across an agent's doxastic possibilities. When you enter a time machine or fall into a coma, your credence might be spread between worlds with very nearby successors and worlds whose successors lie in the distant future, or even in the past.

I do not assume a clear pre-theoretic understanding of the concept of a next world. Rather, I assume that we are interested in the dynamics of belief across certain trajectories, and that these trajectories can be modeled as discrete sequences of worlds, with evidence arriving at various precise points. Any such model determines a successor relation that can be plugged into the amended form of conditioning. It is mathematically routine to relax the modeling assumptions so as to allow for continuous trajectories with a continuous stream of evidence. However, the added mathematical complexity would only obscure the issues I want to discuss. (It may also be worth remembering that the actual world does not seem to contain any genuinely continuous processes.) In practice, the discreteness assumptions are generally harmless. As we will see, when we consider particular agents in particular scenarios, it will usually be easy to construct a discrete model of the relevant update process. It doesn't even matter if we choose the "next world" to be a full day in the future, as long as the agent doesn't receive any relevant evidence in between.

Shifted conditioning, as presented above, does not work if worlds can have multiple successors. Consider the story of Fred. Here the Sunday worlds lie on a branching trajectory that continues with one branch to Monday and with another to Tuesday. On some views in the metaphysics of personal identity, Fred *is* the corresponding branching object, comprising Fred's pre-fission stages together with all stages of his "successors"

(see [Perry 1972]). But again, we can ask what belief transitions would be sensible across this trajectory independently of whether it constitutes the life of a person.

The problem with the above formulation of shifted conditioning is that the successor of a world always inherits the full probability of its ancestor. If there are several successors, the shifted probabilities therefore don't add up to 1. To fix this, I suggest that the probability of a world with multiple successors should be divided among its successors: you can't inherit more than you own.

In Fred's case, the Sunday credence is split between two kinds of worlds: heads worlds with only a Monday successor, and tails worlds with both a Monday and a Tuesday successor. The new shifting process now takes every tails world and assigns *half* its probability to each of its successors. For heads worlds, the full probability is moved to their successors, as before. The effect is that Fred's successors give credence 3/4 to being on Monday and 1/4 to being on Tuesday.

More generally, let $\mathsf{S}w$ be the set of siblings of w, i.e. the set of worlds that are successors of a world of which w is a successor. Let $\#\mathsf{S}w$ be the number of siblings of w. Then redefine the shifting operation so that

$$P^{\succ}(w) = P(\succ w)/\#\mathsf{S}w.$$

That is, the shifted probability of any world w equals the unshifted probability of w's predecessor divided by the number of w's siblings (including w itself). The shifted probability of a proposition is the sum of the shifted probability of its members. As before, the new credence is the shifted previous credence conditioned on the new evidence: $P_2(A) = {P_1}^{\succ}(A/E_2)$.

This essentially completes my proposal. In the next section, I will generalise the account to allow for unequal shifting where some successors inherit more than others. In section 4, I will discuss a rather different approach to fission according to which worlds never have more than one successor. It will turn out that this makes little difference to any concrete application. Some arguments in support of the present proposal will be given in section 6, after I have addressed a possible objection in section 5.

3 Transition Probabilities

[Parfit 1984] has convinced many philosophers that survival comes in degrees. One might similarly argue there are degrees of epistemic successorhood. If w_1 is more of a successor of v than w_2, then arguably more of v's probability should be shifted to w_1. Allowing for unequal inheritance is also crucial in Everettian quantum mechanics, where the redistribution of credence should reflect the quantum mechanical amplitudes of the relevant branches (for reasons reviewed in [Greaves 2007]).

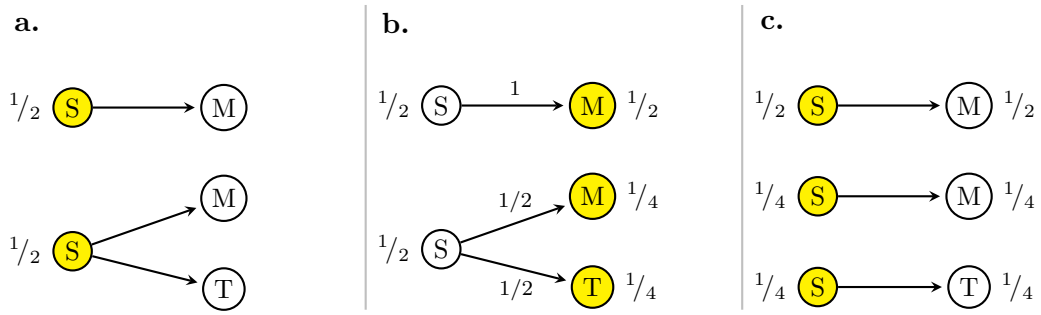

Figure 1: **shifting probabilities in branching worlds.** a. Fred is undecided between two possibilities: either he is about to be teleported to Monday, or he is about to be teleported to both Monday and Tuesday. b. Fred's credence is shifted to the successor points, weighted by the transition probabilities. c. In an Ockhamist framework, the second possibility – being teleported to both Monday and Tuesday – is treated as two distinct possibilities from the outset.

To achieve this kind of generality, we can replace the successor relation by a family of *transition probabilities*, i.e. by a function τ that assigns to each world v a probability distribution τ_v over the space of worlds. $\tau_v(w)$ is the degree to which w is a successor of v. In the previously proposal, $\tau_v(w)$ was fixed as $1/\#\mathsf{S}w$ if w succeeds v, otherwise 0. Now we also allow for other transition probabilities. I assume, however, that $\tau_v(w)$ is 0 unless v and w are worldmates.

Given a transition function τ, the shifted probability $P^{\succ}$ is defined as the expectation of the relevant transition probabilities:

$$P^{\succ}(w) = \sum_{v \in W} P(v)\pi_v(w).$$

That is, to compute the shifted probability of w, we add up the probability of all worlds v that are connected to w, multiplied by the strength of the connection. Once again the final probability is $P_2(A) = P_1^{\succ}(A/E_2)$.

Let's apply this to Fred. Assume the transition probabilities between *Tails & Sunday* worlds and the corresponding Monday and Tuesday worlds are 1/2. To compute the shifted probability of, say, *Tails & Monday*, we add up the old probability of all worlds with links into *Tails & Monday* (i.e. of all *Tails & Sunday* worlds), weighted by the strength of those links (i.e. by 1/2). Since the old probability of *Tails & Sunday* was 1/2, the shifted probability of *Tails & Monday* is 1/4. Figure 1.a–b illustrates the basic idea.

A convenient technique to compute an update goes as follows. Draw a table in which the rows correspond to the relevant old possibilities and the columns to the new ones. (The possibilities will typically not be individual worlds.) Write the previous probability of each row on its side. In each cell of the table, write the probability of transitioning

from the respective old possibility to the new one. More precisely, suppose A is the old possibility and B the new one. Then evaluate, by the light of the previous probability function, the probability of arriving at a B world when propagated forward with the transition probabilities τ, *given* that the present point is an A world. (More precisely still, the value to put in the cell is $\sum_{w \in W} \sum_{v \in B} P(w/A) \times \tau_w(v)$.) Now add up the values in each cell multiplied by the corresponding row probability. The result is the shifted probability distribution, which gets conditioned on the new evidence. It sounds more complicated than it really is. Here is the table for Fred.

		H Mon	T Mon	T Tue
1/2	H Sun	1	0	0
1/2	T Sun	0	1/2	1/2
Shifting:		*1/2*	*1/4*	*1/4*
Conditioning:		*1/2*	*1/4*	*1/4*

The *Heads & Sunday* worlds are only succeeded by *Heads & Monday*, hence the '1 0 0' in the top row. The *Tails & Sunday* worlds, on the other hand, are succeeded by *Tails & Monday* and *Tails & Tuesday* worlds, and the probability of arriving at either sort of world from *Tails & Sunday* is 1/2. The conditioning step is idle here, assuming that nothing relevant is learnt when Fred's successors wake up.

Unlike the version in the previous section, the new version of shifted conditioning can also be applied to cases with infinitely many successors, since the successors no longer have to be given equal probability. We could also make room for *terminal* worlds with no successors, by allowing $\tau_v(w)$ to be 0 for all worlds w. ${P_1}^{\succ}$ would then sometimes fail to be a probability function, but P_2 would always be one. I will instead assume that intuitively terminal worlds are in fact succeeded by arbitrary worlds that are excluded by the new evidence. The two options are equivalent, but the second is technically a little more convenient.[4]

Let me emphasise again that transition probabilities are not some mysterious new kind of probability. The transition probability between a pair of worlds v and w simply reflects the fraction of v's credence that should move to w during shifting. In easy cases, where every world has a unique successor, all transition probabilities are either 0 or 1. (In the next section, we will meet a view on which this covers all cases.) If v has multiple successors, we have to decide how probability should be divided among them when the

4 What about fusion cases, in which a world has several predecessors? The current statement of shifted conditioning allows for such cases, but it will often be unsatisfiable, unless all predecessors have the very same beliefs. If the predecessors have different beliefs, an obvious strategy would be to use a mixture of the their probabilities in place of P_1, as suggested in [Meacham 2010b]. On the other hand, this does not make optimal use of the available information: if one predecessor has found out that $A \vee B$ and another that $\neg A$, why not let the successor know that B (assuming these propositions don't change their truth-values)? For the present paper, I will ignore the possibility of fusion.

agent updates their belief. If you think credence should be divided evenly among the successors, then $\tau_v(w) = 1/n$, where n is the number of v successors (or 0 if w is not among them). If v is an Everett world and you think the redistribution of probabilities should follow the quantum mechanical amplitudes, then $\tau_v(w)$ is the squared modulus of the amplitude of the w branch diverging from v. Of course saying so doesn't make it so. A major challenge for Everettian quantum mechanics, from the present perspective, is to justify these transition probabilities. I do not want to enter into this debate here, nor do I want to settle every other conceivable trouble case. The basic form of the update process remains the same no matter how the transition probabilities are filled in.

4 Ockhamism

Suppose for a moment that Fred knows that he is about to be teletransported to both Monday and Tuesday. Would it nevertheless make sense for him to wonder where he will wake up? I have implicitly assumed that it would not. Uncertainty requires multiple possibilities. For Fred to be uncertain about where he is going to wake up, his doxastically possible worlds would have to divide into worlds where he wakes up only on Monday and worlds where he wakes up only on Tuesday. But then Fred would have misunderstood his situation. Perhaps he thinks he has an immaterial soul that will determinately travel to either Monday or Tuesday. Fred's actual situation is not a world where he wakes up only on Monday, nor a world where he wakes up only on Tuesday. If Fred is aware of the relevant facts, he therefore can't be uncertain about where he will be.

Some philosophers disagree and claim that Fred could meaningfully wonder whether he will be on Monday or on Tuesday. [Ninan 2009] supports this by an appeal to imagination. Couldn't Fred *imagine* waking up on Monday and not on Tuesday? He surely could. *Waking up on Monday* is an ordinary proposition that is true, for example, at Fred's Monday successor. But what follows from the fact that Fred can imagine this proposition? I do not see how this makes it any more plausible that Fred can be uncertain on Sunday about whether he will wake up on Monday.

Another argument in support of pre-fission uncertainty takes off from a certain metaphysical view on personal identity. According to [Lewis 1976], a situation like Fred's really involves two persons, one of which wakes up on Monday and the other on Tuesday. Call these two persons Fred_M and Fred_T. Before the fission, Fred_M and Fred_T are co-located: they occupy the exact same place at the same time. But then we can distinguish two Sunday possibilities: *being* $\textit{Fred}_M$ and *being* $\textit{Fred}_T$. The first is true for Fred_M, the second for Fred_T. Each of these possibilities has a unique, non-branching successor. The co-located Freds on Sunday can be uncertain about where they will be tomorrow by not knowing which Fred they are today.

This line of thought has recently been explored in the context of the Everett interpre-

tation (see e.g. [Saunders 1998], [Saunders and Wallace 2008], [Lewis 2007b], [Tappenden 2008]). The discussion is obscured not only by the metaphysics of personal identity, but also by the unfortunate choice of sentences, rather than propositions, as the bearers of probability. A central topic in the debate is therefore the interpretation of sentences like "I am going to be on Monday", when uttered by Fred on Sunday: is the sentence true under these conditions? If there are two Freds on Sunday, are there also two utterances? And who, then, is the referent of "I"?

From the present perspective, it does not matter how we answer these questions. What matters is whether Fred's credence is divided between possibilities with only a Monday successor and other possibilities with only a Tuesday successor. What, if anything, he thinks about the words "I am going to be on Monday" is besides the point. For all I've said, Fred may not speak English at all, nor any other language. If he speaks English, he might be reluctant to flatly accept or deny "I am going to be on Monday". But this would not show that he is uncertain about the relevant facts. Perhaps the sentence is semantically indeterminate, because the conventions of English don't settle how to talk about one's fissioning future. The kind of reluctance we display towards indeterminate sentences, however, must be sharply distinguished from intermediate credence. Fred might also be genuinely uncertain about whether the sentence is true or false. But again, this merely linguistic uncertainty is besides the point.

There is an old position in tense logic according to which statements about the future in a world with branching time can only be evaluated relative to a particular branch. [Prior 1967] called this view *Ockhamism*. Since every branch determines a unique future, sentences like "there will be a sea battle" have a determinate truth-value at every evaluation point, even if there is a sea battle only on some branch of the future. Similarly, one could say that Fred's utterance of "I will be on Monday" must be evaluated relative to a linear subset of Fred's trajectory. The sentence would be true relative to one branch and false relative to another. Returning to matters of belief, let *doxastic Ockhamism* be the view that every maximally specific possibility in an agent's belief space has a determinate, linear future. Distinguishing Fred_M and Fred_T as alternative Sunday possibilities may achieve this in the scenario of Fred. But we don't need to accept Lewis's implausible metaphysics of personal identity. Whatever we say about persons, we can ask whether the possibilities in Fred's belief space should be modeled by "disambiguating" branching structures or not.[5]

5 For the Everett interpretation, the Lewisian view of persons is not of much help anyway since the universe will keep branching after any particular agent's death. This is why [Saunders 2010] and [Wilson 2011] have suggested moving to a more general kind of Ockhamism.

If we keep belief update separate from questions of personal identity, we must not be led astray by propositions about personal future. Suppose again we decide that persons cannot survive episodes of fission. Then $\succ$ *alive* is true for Fred on Sunday, although Fred cannot truly believe that he (qua person) will be alive. So we must distinguish $\succ A$ from propositions about what is going to be true for

Formally, a non-Ockhamist model is easily converted into an Ockhamist model. We can simply pair each world with a determinate branch. In the Ockhamist model, every world is guaranteed to have at most one successor, so we can stick to the original form of shifted conditioning in section 2: the later revisions to account for cases of fission are redundant; all transition probabilities are either 0 or 1.

Let's use the Ockhamist framework to model the story of Fred, returning to the original case where he does not know what will happen. We now start with *three* possibilities on Sunday: a heads possibility leading to Monday, a tails possibility leading to Monday, and a tails possibility leading to Tuesday. How is Fred's Sunday credence divided between these alternatives? Since the coin is fair, he should give equal credence to heads and tails. Within the tails possibilities, he should perhaps give equal credence to the possibility leading to Monday and the possibility leading to Tuesday. Applying shifted conditioning then yields the same result as before (see figure 1.c).

In general, where we previously saw a single possibility with several futures, we now see several possibilities with unique futures. Since these possibilities are at present indistinguishable by the agent, the questions arises how rational credence should be divided among them. All the options for transition probabilities then return. Following the simple proposal in section 2, for example, we could say that credence should always be evenly distributed among such worlds. Alternatively, we could say that for Everett worlds the distribution should accord with the quantum mechanical amplitudes. And so on. For any given answer, the result of applying shifted conditioning under Ockhamism will be exactly the same as it is without Ockhamism, using the corresponding choice of transition probabilities.

The upshot is that it matters little whether we use an Ockhamist or a non-Ockhamist model. In my view, Ockhamist models distort the doxastic situation of agents in expectation of fission, by postulating uncertainty where there is nothing to be uncertain about. But it is reassuring that in the framework of shifted conditioning, this somewhat esoteric question makes no real difference. There is, however, a technical advantage of Ockhamist models. Ockhamism retains the equation $P^{\succ}(A) = P(\succ A)$. This can be very convenient. I will exploit it in section 6 to streamline some arguments in support of shifted conditioning. But first I want to respond to a possible objection.

---

the relevant person. Notice also that Fred (qua person) has the property of *not existing at* t_2, where t_2 is a time after the fission. Hence the relevant branching trajectory is a "temporally incoherent" in the sense that any individual whose history is this trajectory would have to exist at t_2 although at t_1 it has the property of not existing at t_2.

5 Manipulated memories, evidence and experts

Standard conditioning has problems with propositions that change their truth-value. Some argue that it also goes wrong in cases where the agent has reason to distrust their memories. Consider the scenario of *Shangri La*, from [Arntzenius 2003]. At t_1, you get to see the outcome of a fair coin toss. Later, at t_2, you walk through a gate which replaces any memories you may have of a tails outcome with (quasi-)memories of heads. In fact, the coin lands heads, so your memories are left intact.

After passing through the gate, you are certain to end up with memory experiences of heads, independently of the actual outcome. A common intuition is that you therefore ought not to be confident at this stage that the coin landed heads. Conditioning disagrees: since all your t_1 credence went to heads worlds, and your new evidence is equally compatible with heads and tails, your t_2 credence should still be concentrated on heads worlds.

I think conditioning gives the right answer here. What might be driving the contrary intuition is the idea that it would be unreasonable for you to trust your memory evidence at t_2. In the terminology of [Pollock 1986], your knowledge of the setup provides an "undercutting defeater" for the testimony of your memory. But conditioning does not say that you should trust your memories. The reason you should believe in heads is not your vivid recollection of seeing this outcome. On the contrary, conditioning says that you should believe in heads because that's what you believed before and your new evidence provides no information one way or the other. It is not your memories that do the work, but your previous beliefs.

Instead of asking how your beliefs should change between t_1 and t_2, we might ask a different question: to what extent does your evidence at t_2 supports the hypothesis that the coin landed heads? The answer to *this* question is plausibly around 1/2. But the two questions must not be conflated. When I ask how your beliefs should evolve, I want to know which new belief state would constitute an optimal update of your previous belief state in the light of the new evidence. The challenge, in general, is to devise an algorithm that takes as input an old probability distribution and an evidence proposition and returns a new probability distribution. On the other hand, when we ask to what extent some hypothesis is supported by such-and-such evidence, it does not matter whether anyone actually has this evidence, nor what they believed in the past. Here, the challenge is to devise an algorithm for an "oracle" that takes as input arbitrary evidence propositions and returns an appropriate probability distribution. Scientific confirmation is arguably a matter of these evidential probabilities, rather than any actual agent's degrees of belief.

Perhaps the two questions are often conflated because we humans generally have evidence in support of whatever we should rationally believe, in the form of suitable

memories or introspective evidence. Some philosophers also subscribe to the doctrine of *evidentialism*, according to which rational belief is only constrained by present evidence (see e.g. [Feldman and Conee 1985]). This might seem to entail that an agent's rational credence at any time should coincide with the evidential probability conditional on their total evidence. I will argue in a moment that this is false: even evidentialists should distinguish evidential probability from rational credence.

Evidentialism is a radical position. It is incompatible with any norm that relates an agent's present beliefs to their earlier or later beliefs. It rules out conditioning and shifted conditioning as well as any other norm that determines the new probabilities in part by the old ones. The previous probabilities could only be used if they are given by the new evidence. In general, the evidentialist counterpart of a norm relating present credence to previous credence takes the form of an *inverse reflection* principle. For shifted conditioning, inverse reflection states that the agent's present probabilities should equal their expectation of the shifted previous probabilities conditional on the new evidence. More strictly,

$$P_2(A/P_{-1}{}^{\succ}(A/E){=}x \;\&\; E) = x,$$

where P_{-1} *non-rigidly* denotes the previous probability. Agents who know that they follow shifted conditioning automatically satisfy this principle.

In the following section, I will give a number of arguments in support of shifted conditioning. These are also arguments against evidentialism. Nevertheless, evidentialists can reap many benefits of shifted conditioning by endorsing the corresponding principle of inverse reflection – perhaps restricted to situations in which the evidence entails that the previous credence was rational. If an agent is fortunate enough to have perfect evidence about their previous beliefs, adherence to inverse reflection has the same effect as updating by shifted conditioning.

In Shangri La, your new evidence is equally compatible with the hypothesis that your previous credence in heads was very high and the hypothesis that it was very low. Inverse reflection therefore does not entail that your new credence in heads should be high. Shifted conditioning, on the other hand, agrees with conditioning: if at t_1 you learn that a coin landed heads, and at t_2 you receive evidence that is equally compatible with heads and tails, then you should keep believing that the coin landed heads. As an answer to the update question, this sounds right to me.

Notice that if you update by conditioning or shifted conditioning, then you know both before and after walking through the gate that your memories are left intact. Compare the situation of an agent who knows that whenever they walk through a particular gate, various false beliefs get injected into their mental state. After having walked through the gate, this agent might be better off, epistemically speaking, to put more trust in their present evidence than in their possibly corrupted previous beliefs. That is, the

best update algorithm under those circumstances might be one that determines the new probabilities as a mixture of the evidential probabilities and the result of shifted conditioning.[6] The less trustworthy the previous beliefs, the more the mixture should be skewed towards the evidential probabilities. But Shangri La is not a scenario of this kind, at least not if you update by conditioning or shifted conditioning. So the present consideration does not support an argument against conditioning.

Return once more to the distinction between evidential probability and rational credence. That the two come apart is clear if rational credence is subject to diachronic norms relating the present beliefs to previous beliefs. It is not so clear for evidentialists, who reject genuinely diachronic constraints on rational belief. However, I have argued that evidentialists should endorse the principle of inverse reflection corresponding to shifted conditioning. I am now going to argue that evidential probabilities, by contrast, do not satisfy this principle.

For evidential probabilities, the principle of inverse reflection says that the probability of a proposition A, conditional on some information E together with the information that the shifted credence in A given E of the agent at the predecessor of the present situation was x, should be x. As above, this should presumably be restricted to situations in which E entails that the previous credence was rational. The principle might then be motivated as follows.

The fact that a rational agent assigns, say, high credence to A indicates that there is evidence out there in support of A. But evidence for evidence for A is evidence for A. On the other hand, we also have to take into account the further information E. It might be that the agent's evidence, combined with our evidence E, would make A very unlikely. So we should consider not the agent's actual credence in A, but their credence in A conditional on E. That is, if E entails that f is a rational agent's probability function, then

$$P(A/f(A/E){=}x \;\&\; E) = x.$$

This is an *expert principle* of the kind discussed e.g. in [Gaifman 1988] and [van Fraassen 1989: 201f.].

In this form, the principle ignores that what is true for the relevant agent (the "expert") may be false at the target situation. To adjust for these differences, we need to know how the two locations are related to one another. If the expert's situation is located 5000 km to the North and 80000 years in the past, then her belief that there are Mammoths nearby may only support the hypothesis that there were Mammoth 80000 years ago at around 5000 km to the North.

6 *Ideally*, the agent would still obey shifted conditioning at all times, rather than the mixed update. This would entail that they never start believing things that are not supported by previous beliefs and current evidence. The mixed rule is a fall-back algorithm for cases where the ideal update is impossible. The mixed rule may also be reasonable for creatures with imperfect memories, like us.

The required transformations are easy if we know that the expert is located at the predecessor of the target situation. Setting aside fission, this means that A is true at the target situation iff $\succ A$ true for the expert. The expert principle can therefore be corrected as follows:

$$P(A/P_{-1}(\succ A/\succ E) = x \;\&\; E) = x.$$

In the absence of fission, this is equivalent to the principle of inverse reflection. If the expert is about to undergo fission, their credence in $\succ A$ given $\succ E$ may be high because they know that E is true at one of their successors and A at another. Obviously this does not support the hypothesis that A is true at the target situation given that E is true. This problem is avoided by using $P_{-1}{}^{\succ}(A/E)$ instead of $P_{-1}(\succ A/\succ E)$ (or, equivalently, by assuming Ockhamism).

So what's wrong? First of all, suppose the evidence entails that the target situation is either on Monday or on Tuesday. As it happens, the situations trace back to a common ancestor, at which the transition probabilities favoured Monday over Tuesday: perhaps the Monday transition had a higher degree of survival, or higher quantum mechanical amplitude. Should this skew the evidential probability towards Monday?

Or consider the following situation. In 2011, Frieda is undecided between being on Earth and being on Twin Earth. Twin Earth, she knows, will undergo planetary fission on New Year's Eve 2012, leaving behind two indistinguishable copies of itself. After the fission, Frieda's evidence is equally compatible with being on Earth, being on Twin Earth 1 and being on Twin Earth 2. Since she updates by shifted conditioning, her credence in being on Earth remains at 1/2. Some years later, Frieda gives birth to a daughter. The daughter, as she grows up, learns that the universe contains three planets indistinguishable from the one she inhabits: Earth, Twin Earth 1 and Twin Earth 2. Later she learns that the two Twin Earths came into existence by planetary fission. This cosmological fact does not seem to shed any light on where she herself might be, and thus should not significantly increase the daughter's credence in being on Earth. Her rational credence in being on Earth may therefore be less than 1/2. Moreover, her credence would be in line with the evidential probabilities: arguably, her total evidence supports being on Earth to a degree less than 1/2. But then it is hard to see why *Frieda*'s total evidence, at the same time, should lend more support to being on Earth. The only relevant difference between Frieda's evidence and her daughter's is that Frieda's evidence entails that the target situation lies on an epistemic trajectory that may itself have undergone fission. But why is this any more relevant than the historical information about planets? Note also that Frieda's daughter may well know that her mother's credence in being on Earth (past and present) is 1/2. For she may know that her mother was undecided between Earth and Twin Earth before the fission and then updated by shifted conditioning. Again this information does not seem to shed any light on where she may be. But if her mother

should not be treated as an expert at the daughter's situation, why would we have to treat her as an expert at situations that lie on the mother's own trajectory?

If this is correct, the scenario of Frieda not only shows that evidential probabilities do no obey inverse reflection, it also provides an interesting case of peer disagreement: Frieda and her daughter would rationally disagree about the probability of them both being on Earth, even after exchanging all the information they have, and even if they started off with the same ultimate priors.

I have assumed that rational agents update their beliefs by shifted conditioning. It is high time to give reasons for this assumption, and against various alternatives that have been proposed.

6 Diachronic norms and the problem of drift

Many philosophers draw a sharp distinction between *de dicto* beliefs with uncentred content and self-locating beliefs concerning the agent's place within the world. A common picture of belief update then combines standard conditioning for uncentred propositions with a new rule for self-locating beliefs, as follows.

Let P_1^* be the previous probability function P_1 restricted to uncentred propositions. For any proposition A, let $\Diamond A$ be the strongest uncentred proposition entailed by A. Given an evidence proposition E_2, define P_2^* as P_1^* conditioned on $\Diamond E_2$. Each uncentred world to which P_2^* assigns positive probability contains at least one centre at which E_2 is true. If it contains exactly one such centre, the agent knows that she must be at this point if she is anywhere in the uncentred world. If E_2 is true at several points, the standard procedure is to use a principle of *self-locating indifference*. This says that all centres within an uncentred world that are compatible with the evidence should have equal probability. One way to define the full posterior probability P_2 is then to evenly divide the P_2^*-probability of each uncentred world among those of its centres where E is true. This leads to halfing in the Sleeping Beauty problem. A more popular alternative, which supports thirding, is to assign the whole P_2^*-probability of each uncentred world to all its E_2-centres and then renormalise the probability distribution. Proposals along these lines are defended e.g. in [Piccione and Rubinstein 1997], [Halpern 2006], [Meacham 2008], [Titelbaum 2008], [Kim 2009] and [Briggs 2010].

These accounts are *hybrids* insofar as they combine a diachronic rule for uncentred propositions with an evidentialist rule for self-locating beliefs: within any uncentred world, the new probabilities are determined only by the new evidence – the previous self-locating beliefs are completely ignored.

A consequence of this is what might be called *the problem of drift*. For a simple illustration, suppose your credence is divided between three uncentred worlds w_1, w_2 and w_3, and you know that you are going to learn either E or E'. E is true at exactly one

point in each of w_1 and w_2; E' is true at exactly one point in w_1 and w_3, but the point in w_1 is different from the point where E is true. Suppose you presently give credence $1/3$ to each of w_1 through w_3. If you then learn E and update by the hybrid rule just described (using either the halfer of the thirder version), your credence in w_1 will increase to $1/2$. Likewise if you learn E'. *No matter what you learn*, your credence in w_1 will go up!

A prominent instance of this phenomenon arises in Everettian quantum mechanics, where, however, it is often misdiagnosed either as a problem for Everettians or as a problem for thirders (see e.g. [Lewis 2007a], [Bradley 2011]). Suppose you are neither absolutely certain that the Everett interpretation is true, nor that it is false. Whenever you toss a coin, the Everett worlds in your belief space contain a branch on which the outcome is heads and a branch on which it is tails. The non-branching worlds in your belief space only contain either heads or tails. According to the hybrid update rules, observing either outcome should increase your credence in the Everett hypothesis. By repeatedly tossing coins, you would become more and more confident in the Everett hypothesis, no matter what outcomes you observe. Clearly something has gone wrong.

Hybrid accounts lead to this kind of probability drift whenever multiple evidence propositions are true within the same uncentred world. The underlying cause is the evidentialist treatment of self-locating beliefs. When you see the coin landing heads, the hybrid accounts let you condition only on the uncentred proposition that the universe contains some point or other at which the coin lands heads. This rules out non-Everett worlds in which the coin lands tails, but it does not rule out any Everett worlds. Self-locating indifference is then used to locate yourself within the remaining worlds. But what you actually learn isn't just a fact about the universe as a whole. You also learn that you are presently looking at a heads result. Conditioning on this information would exclude tails possibilities in Everett worlds just as much as it excludes tails possiblities in non-Everett worlds. This is what happens if you follow shifted conditioning. Since the shifting step also leaves the probability of uncentred propositions untouched, the problem of drift doesn't arise.

We may turn this into a general constraint on update models.

> *Dynamic stability.* If a proposition is certain not to change its truth-value, then rationality should not require its probability to increase independently of what evidence is received.

Dynamic stability is closely related to various "Minimal Revision" constraints. For example, [Teller 1973] discusses the condition that if two propositions initially have equal probability, and both entail the new evidence E, then they should still have equal probability after updating on E. This needs to adjusted to allow for centred evidence. In

[Schwarz 2011] I show that shifted conditioning satisfies the adjusted constraint, while hybrid accounts do not.

The problem of drift can also be turned into a diachronic Dutch book. Return to the "three worlds" example from above. Here you initially regard as fair a bet that pays \$-4 in case of w_1 and \$2 otherwise; after updating by the hybrid rules, you would regard as fair a bet that pays \$3 in case of w_1 and \$-3 otherwise – no matter what you have learned. Taken together, these two bets amount to a sure loss of \$1.[7]

Let *dynamic coherence* be the requirement that an update rule should not be exploitable in this way by a Dutch book. Hybrid accounts fail the requirement of dynamic coherence. In [Schwarz 2011] I showed that every alternative to shifted conditioning is dynamically incoherent. The converse, that shifted conditioning itself is dynamically coherent, can be proved as follows, adapting an argument in [Skyrms 1987].

Suppose there is a diachronic Dutch book against an agent who obeys shifted conditioning. A diachronic Dutch book consists of a (finite) set of earlier bets together with a mapping from the members of some evidence partition to (finite) sets of later bets. Consider one of these later bets, to be placed if the new evidence is E_i. Let's say the bet pays $\$X$ in case of A and $\$Y$ otherwise. Since it is part of a diachronic Dutch book, we know that the agent would take it to have positive payoff after learning E_i: $P_2(A)\$X + P_2(\neg A)\$Y \geq \$0$, where P_2 is P_1 updated on the information E_i. Given Ockhamism $P_2(A) = P_1^{\succ}(A/E_i) = P_1(\succ A/\succ E_i)$, and so $P_1(\succ A/\succ E_i)\$X + P_1(\succ \neg A/\succ E_i)\$Y \geq \$0$. This means that the agent should already accept at the earlier time a bet that pays $\$X$ if $(\succ A)$ & $(\succ E_i)$, $\$Y$ if $(\succ \neg A)$ & $(\succ E_i)$ and \$0 if $\neg\succ E_i$. Since $\succ A$ and $\succ E_i$ are true at the earlier time iff A and E_i are true at the later time (respectively), the actual payoff is guaranteed to be the same for the original later bet and the converted earlier bet. Substituting each of the later bets by a corresponding earlier bet, and combining these bets with the original earlier bets therefore yields a synchronic Dutch book against the agent at the earlier time. But [Kemeny 1955] proved that if an agent's probabilities respect the probability calculus, then they are immune to (finite) synchronic Dutch books. It follows that an agent who obeys the probability calculus and updates by shifted conditioning is also

7 [Briggs 2010] presents a purported proof that the (thirder version of the) hybrid update rule is immune to diachronic Dutch books. (Her rule differs slightly from the rule presented above, but it is equally vulnerable to the Dutch book just given.) Her argument is that if there *were* a Dutch book B for an agent who follows the hybrid update rule, then we could convert B into a Dutch book B' for an imaginary agent with only uncentred beliefs who updates by standard conditioning; but the latter is impossible by a result in [Skyrms 1987] (falsely attribute by Briggs to [Teller 1973]). Applying Briggs's recipe to my Dutch book in the "three worlds" scenario, B' is identical to B. Since the imaginary agent initially assigns equal probability to all three worlds, she regards the first bet as fair. After conditioning on either $\Diamond E$ or $\Diamond E'$, she will regard the second bet as fair. However, *pace* Briggs, this pair of bets does not constitute a Dutch book against the imaginary agent. The problem is that $\Diamond E$ and $\Diamond E'$, unlike E and E', are not mutually exclusive: they are both true at w_1.

immune to diachronic Dutch books.[8]

A third norm that speaks in favour shifted conditioning is *reflectivity*. [van Fraassen 1984] argues that one's rational credence in an uncentred proposition A should equal the expectation of one's future credence in A: $P_1(A) = \sum_x P_1(P_2(A)=x)\ x$. To bracket situations in which you expect your future self to have undergone cognitive mishaps, I will focus on the following, more restricted version:

> *Reflectivity.* If you know that you are about to update your beliefs in a rational way, and A is certain not to change its truth-value, then your present credence in A should equal your expectation of your future credence in A.

Unlike the requirements of dynamic coherence and stability, reflectivity is a synchronic condition, since it does not involve the actual future probability. It is easy to show that shifted conditioning satisfies this requirement while hybrid accounts, for example, do not. Once again, the problem is evident in a case of drift. If you know that you follow a hybrid rule under which A becomes more and more probable no matter what you observe, then your present credence in A will be lower than your expected future credence.[9]

Drift can also affect centred propositions. [Meacham 2010b] defends a version of shifted conditioning on which $P_2(A) = P_1^M(A/E_2)$, where the shifting transformation M is

8 If bets are offered both before and after an episode of fission, how shall we evaluate the total payoff for the involved agents? The two fission products are clearly different persons, so their gains or losses arguably shouldn't be added together into the total payoff for any single individual. What about the pre-fission payoff? By assuming Ockhamism, I effectively let both post-fission persons count the pre-fission payoff as their own. I will return to these matters in the discussion of Sleeping Beauty in section 8.

9 In cases involving fission, we have to be careful about the right interpretation of "your future credence". If Fred's Monday successor finds out that he is on Monday, his credence in heads increases to 2/3. Suppose on Sunday, Fred knew that his Monday successor would find out where he is. Then Fred's credence in heads is 1/2 although he knows, in a sense, that his future credence will be 2/3. But here the relevant "future credence" is the credence of Fred's Monday successor. Fred's Tuesday successor, if he has one, will give credence 0 to heads after discovering where he is. Since Fred is assigns credence 1/2 to having a Tuesday successor, a sensible way to compute the "expectation of his future credence" is to take expectation of the average of all the future credences (weighted by the transition probabilities), which is $1/2 \times 2/3 + 1/2 \times (2/3 + 0)/2 = 1/2$.

defined by[10]

$$P^M(w) = P(\succ w)\,\frac{P(\diamond w)}{\sum_{v\in\diamond w} P(\succ v)}.$$

Now suppose a certain universe contains three planets where you might be; call them Earth 1, Earth 2 and Earth 3. On Earth 1, the coin you are about to toss lands heads. On Earth 2, it lands tails. On Earth 3, you will undergo fission, with one of your successors witnessing heads, the other tails. If your initial probability in each of the three locations is 1/3, then Meacham's shifted probability assigns $1/3 \times \frac{1}{4/3} = 1/4$ to each of the four successor locations. The probability of Earth 3 thereby increases to 1/2. Conditioning either on heads or on tails leaves it at 1/2. Just like on the hybrid proposals, probability drifts from the non-fission hypothesis to the fission hypothesis.

In my own formulation of shifted conditioning, it does not matter whether two alternatives happen to be worldmates or not. When you see the coin land heads (for example), the update goes like this.

		H E1	T E2	H E3	T E3
1/3	E1	1	0	0	0
1/3	E2	0	1	0	0
1/3	E3	0	0	1/2	1/2
Shifting:		*1/3*	*1/3*	*1/6*	*1/6*
Conditioning:		*2/3*	*0*	*1/3*	*0*

The probability of the branching hypothesis remains at 1/3.

Observe that *Heads & Earth 1* and *Heads & Earth 3* are located in the same uncentred world, but the former is now twice as probable as the latter. This illustrates that shifted conditioning is incompatible with another popular constraint on rationality: the principle of self-locating indifference.[11] We have already seen this in the Frieda example from the previous section: Frieda's credence in being on Earth is 1/2, although her evidence is equally compatible with Earth, Twin Earth 1 and Twin Earth 2.

10 This is a corrected reformulation of Meacham's "Local Predecessor Conditionalization". ([Meacham 2010b] also discusses a "Global" version that leads to drift even among uncentred propositions.) Meacham himself defines

$$P_2(A) = \sum_{w\in A} P_1(\succ w/\succ E)\frac{P_1(\diamond\succ w/\succ E)}{\sum_{v\in\diamond w} P_1(\succ v/\succ E)}.$$

By summing over all worlds that are successors of worlds that have E-worlds as successors, this lets worlds have positive probability that are incompatible with the new evidence. My own formulation avoids this problem, and matches Meacham's informal presentation of his rule. Meacham (p.c.) agrees.

11 Thanks to Chris Meacham for pushing me here. Meacham's version of shifted conditioning respects self-locating indifference in this example, but not in others, such as the "virtual reality" example below.

This violation of self-locating indifference is an inevitable consequence of avoiding drift. In order to avoid drift, the probability of *Earth 3* must remain at 1/3. On the other hand, the only open possibilities after updating on the new evidence are *Heads & Earth 1* and *Heads & Earth 3*. So the two open centres in the relevant *Heads* worlds cannot have equal probability. We have a direct clash between dynamic stability, coherence and reflectivity on the one hand, and self-locating indifference on the other.

To my mind, this provides a strong reason against self-locating indifference as a general constraint on rational belief. We can still allow self-locating indifference as a constraint on *evidential* probabilities. In Frieda's situation, I've argued that the three possible locations may have equal evidential probability. We can also allow self-locating indifference as a constraint on *ultimate priors*. That is, shifted conditioning is compatible with saying that any rational epistemic trajectory must begin with a probability function that obeys self-locating indifference.[12]

It is worth noting that the unrestricted principle of self-locating indifference is not as harmless and intuitive as it may look in selected applications like the Sleeping Beauty problem. Suppose on some distant planet there is a civilisation of aliens with very advanced "virtual reality" technologies. At some point in the year 3011, a member of this civilisation will spend half an hour in a virtual reality device where, by sheer chance, she will have experiences indistinguishable from the ones you yourself will have tomorrow at noon. Suppose for some reason you know that this is the case. Assuming that evidence supervenes on experiences, self-locating indifference would require you to become completely undecided tomorrow at noon about whether you live in the 21st century on Earth or at a much later time on a distant planet – despite the fact that you were certain about being on Earth only moments before and that no surprising information has arrived at noon. Far from being a requirement of rationality, this change of belief would strike me as highly irrational. The probability of a skeptical hypothesis should not dramatically increase in response to completely unsurprising information.[13]

12 This is so if we restrict ourselves to finitely many worlds. It is less clear what happens in infinitary cases, mostly because it is unclear how the indifference principle should then be spelled out. For example, if an uncentred world contains denumerably many positions compatible with the evidence, the only uniform distribution would assign all of them probability zero, which leads to violations of countable additivity.

13 A further drawback of self-locating indifference is that it attaches great weight to the choice between Ockhamist and non-Ockhamist models: combined with Ockhamism, self-locating indifference would require your initial credence in *Earth 3* to be 1/2 instead of 1/3, as Ockhamism divides *Earth 3* into two more specific possibilities.

7 Chance, credence and inadmissible information

I have left out some details about the story of Fred. First of all, I didn't tell you that Tuesday (the moon) is quite a bit further away than Monday. Second, the coin that decides whether Fred gets teleported to Tuesday is actually tossed by Fred's successor on Monday. Recall that the coin toss decides whether the signal to Tuesday gets destroyed. This is possible because by the time Fred has been recreated on Monday, the signal to Tuesday is still on its way.

When Fred's Monday successor tosses the coin, he knows that he is on Monday, for the Tuesday successor doesn't get to toss a coin. If his initial credence was divided 1/2, 1/4, 1/4 between *Heads & Monday*, *Tails & Monday* and *Tails & Tuesday*, his new credence after learning that it is Monday should be divided *2/3*, *1/3* between the first two possibilities. A single update matrix illustrates the whole process:

		Mon,H	Mon,T	Tue,T
1/2	Sun,H	1	0	0
1/2	Sun,T	0	1/2	1/2
Shifting:		*1/2*	*1/4*	*1/4*
Conditioning:		*2/3*	*1/3*	*0*

I assume that on Sunday, when Fred already knew that his Monday successor will toss a fair coin, he gave equal credence to heads and to tails. The result of shifting corresponds to the state of Fred's successor immediately after awakening, when he has not yet discovered where he is.

So Fred's Monday successor (let's call him Fred) should assign probability 2/3 to the assumption that the fair coin he is about to toss will land heads. This is remarkable because agents who know that the objective chance of a future event is x should usually believe to degree x that the event will occur.

Is this a problem for shifted conditioning? I don't think so. Before he learned that he is on Monday, Fred must have somehow divided his credence between *Heads & Monday*, *Tails & Monday* and *Tails & Tuesday*. Learning that he is on Monday rules out one of the tails possibilities and therefore increases the probability of heads. How could the result of this increase have been 1/2? Given that the probability of heads was 1/2 on Sunday, this would lead right back into the problems of drift.

[Lewis 2001] argued that the Sleeping Beauty problem provides an analogous counterexample to the usual connection between chance and rational credence. According to Lewis, Sleeping Beauty, like Fred, should assign credence 2/3 to heads after learning that it is Monday, even if the coin is only tossed on Monday night. In the subsequent discussion, almost everyone concluded that Lewis was wrong about this. As we'll see in the next section, I agree with this verdict, at least on the most widespread interpretation

of the Sleeping Beauty problem. Nevertheless, what Lewis said about Sleeping Beauty is right about Fissioning Fred.

This reveals some important lessons about the link between chance and rational credence. A classic formulation of this link, Lewis's *Principal Principle*, states that

(PP) $\quad P_0(A/Ch_t(A)=x \;\&\; E) = x, \text{ if defined,}$

where P_0 is a rational prior credence function, $Ch_t(A)=x$ is the proposition that the objective chance of A at time t is x, and E is any "admissible" further information. Admissible information, Lewis explains, is "the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of those outcomes" [Lewis 1980: 92]. The restriction to prior credence functions excludes cases in which the agent has already gained inadmissible information.

Inadmissible information is easy to come by if t is in the past: after you've seen the coin land heads, your credence in heads conditional on the initial chance being 1/2 will not be 1/2, because you have received the inadmissible information about the actual outcome. On the other hand, suppose we only consider present chances, replacing $Ch_t(A) = x$ in (PP) by $Ch_{now}(A)=x$.[14] After the coin has landed, the chance of this toss coming up heads is arguably 1. Observing the outcome therefore does not provide you with inadmissible evidence with respect to the present chance of heads. Is it still possible for agents to have inadmissible information? If not, then all rational credence functions P, prior or not, should satisfy

(PP*) $\quad P(A/Ch_{now}(A)=x) = x, \text{ if defined.}$

The stock argument against (PP*) involves crystal balls. Suppose a crystal ball tells you that the fair coin you are about to toss will land heads. The argument is that this should raise your credence in heads, even if you know that the objective chance is 1/2. Now, the pronouncement of the crystal ball shouldn't affect your credence if you think it is completely unreliable. If your credence in heads goes up, you presumably believe that there is some kind of connection between the prediction and the outcome: perhaps the laws of nature guarantee that whatever the crystal ball predicts will actually come about, or that it will come about with reasonably high objective chance. (This doesn't mean that the prediction causes the outcome. Causation may well go in the other direction.) But if the world is like this, then the chance of heads, after the prediction, is no longer 1/2. So the argument from crystal balls doesn't seem to work.

Nevertheless, (PP*) is false, as illustrated by the story of Fred. Unlike in the crystal ball scenario, there is no reason to doubt that at the time when Fred is about to toss

14 I am not suggesting that we should give up the original version of the Principle in favour of this version. As [Meacham 2010a] argues, rational belief (and, one should add, degree of confirmation) are also guided by information about past chances.

the coin, the objective chance of heads is 1/2, and that Fred could know this. Yet his rational credence is 2/3.

Is this because Fred has inadmissible evidence? Following [Lewis 2001], one might argue that the information that he is on Monday is inadmissible, since it rules out *Tails & Tuesday* and thereby reveals something about the outcome of the coin toss over and above what is revealed by the chances. But is this correct? Let t_1 be the time at which Fred wakes up, t_2 the time at which he learns that he is on Monday, and t_3 the time at which the Tuesday successor wakes up. At t_1, Fred does not know whether it is t_1 or t_3. He does know, however, that if it is t_3, then the coin has already landed tails and so the present chance of tails is 1 (let's assume). The information that he is on Monday thus does not reveal anything about the chances at t_1, but it does reveal something about the *present* chances: it reveals that the present chances are the chances at t_1. This suggests that the *Monday* information is inadmissible with respect to $Ch_{t_1}(Heads) = 1/2$, but not with respect to $Ch_{now}(Heads) = 1/2$. On the other hand, at t_1, $P(Heads/Ch_{now}(Heads) = 1/2) = 2/3$, so it looks like Fred already has inadmissible evidence with respect to $Ch_{now}(Heads) = 1/2$.

However, the fact that an agent fails to satisfy (PP*) does not entail that they have inadmissible information in the sense relevant to (PP), unless the agent's credence results from a rational prior probability by standard conditioning on their present evidence – which it shouldn't. To see whether Fred's evidence is inadmissible in the sense relevant to (PP), let E_1 and E_2 be Fred's total evidence at t_1 and t_2, respectively. E_1 is compatible with three possibilities: *Heads & Monday*, *Tails & Monday*, *Tails & Tuesday*; E_2 is only compatible with the former two. It is very plausible that for any rational prior probability P_0, $P_0(Tuesday/E_1) > 0$, and so $P_0(Heads/E_1) < P_0(Heads/E_2)$. But both E_1 and E_2 entail that $Ch_{t_1}(Heads) = 1/2$. So either E_1 or E_2 must contain inadmissible information with respect to $Ch_{t_1}(Heads) = 1/2$. On the other hand, E_1 entails that $Ch_{now}(Heads) = 1/2$ iff *now* is t_1. So $P_0(Heads/Ch_{now}(Heads) = 1/2 \;\&\; E_1) = P_0(Heads/Ch_{now}(Heads) = 1/2 \;\&\; E_2)$. Moreover, one might argue that these should have a value of 1/2, and correspondingly that $P_0(Heads/E_1) = 1/3$. Then there would be no inadmissible evidence with respect to $Ch_{now}(Heads) = 1/2$.

A second lesson from the story of Fred is therefore that an agent's rational credence in a chancy proposition may fail to match their expectation of the present chance, even though they do not have inadmissible evidence about the present chance.

Here is a third lesson. Several authors have argued that (PP) should be reformulated to get rid of the admissibility restriction. Let $Ch_t = f$ be the proposition that the probability function f plays the chance role at time t. Assuming that information about chance is admissible, and setting aside worries about infinitely many possible chance functions,

(PP) seems to entail that for arbitrary A and E,

$$P_0(A \,\&\, E \,/\, Ch_t = f) = f(A \,\&\, E), \text{and}$$
$$P_0(E \,/\, Ch_t = f) = f(E).$$

By the probability calculus, it follows that

(PP**) $P_0(A \,/\, Ch_t = f \,\&\, E) = f(A \,\&\, E)/f(E) = f(A/E)$.

Thus if we use *conditional chances* on the right-hand side of the Principle, we can drop the admissibility restriction. Less formally, imagine that objective chance is an expert whose judgements you trust. If you have relevant information which the expert may lack, then you should trust not their actual judgment but their judgment conditional on the further information. This is what (PP**) says (see [Hall 2004]).

However, we've just seen that in the story of Fred, either E_1 or E_2 is inadmissible with respect to $Ch_{t_1}(\textit{Heads}) = 1/2$. More specifically, given the information in E_1, *Monday* rules out *Tails & Tuesday* and thereby raises the probability of *Heads*, without indicating that the t_1 chance of heads is anything but 1/2. Following (PP**), we should therefore consider not the unconditional chance of heads, but the chance of heads conditional on the further information in E_2. But arguably there is such a thing as the chance of heads conditional on the centred information *being on Monday*. (Quantum mechanics certainly doesn't specify any such chance.) So we can't condition the chances on the evidence to get rid of the admissibility restriction.

8 Three answers to the Sleeping Beauty problem

Finally, let's have a quick look at Sleeping Beauty. Recall that on Sunday (the day, not the planet), Beauty learns that if a certain fair coin toss results in tails, then all her memories of Monday will be erased before her awakening on Tuesday. If the coin lands heads, her memory isn't erased but she is made to sleep all through Tuesday. In fact the coin lands tails, but Beauty doesn't know this. How should her beliefs change between Sunday night and Monday morning?

Let's start by considering Beauty's evidence as she wakes up. First of all, she remembers being informed about the experimental setup, and we can assume that she still trusts this information. In addition, her memories suggest to her that she has not been awake since she went to sleep on Sunday. Given her knowledge of the setup, however, it would be unreasonable for her to trust these memories and infer that it must be Monday: Beauty knows that if the coin landed tails, then she would have the exact same memories on Tuesday.

What is the probability that the coin landed heads, given Beauty's evidence? Various considerations suggest that it should be 1/3, or at least less than 1/2. One way to

motivate this is to take the evidence piece by piece (see [Horgan 2004]). Presumably the information about the setup, by itself, is neutral between the four combinations of heads and tails with Monday and Tuesday. Beauty's memory evidence then rules out *Heads & Tuesday* (as well as any possibility on days other than Monday and Tuesday). The resulting probability for heads should therefore be around 1/3.[15]

But this only answers the evidential question. What about the update question? How would an optimal algorithm update Beauty's Sunday credence in light of her evidence on Monday morning? Shifted conditioning provides the answer.

		H Mon	T Mon	T Tue
1/2	H Sun	1	0	0
1/2	T Sun	0	1	0
Shifting:		*1/2*	*1/2*	*0*
Conditioning:		*1/2*	*1/2*	*0*

Since Beauty's Sunday state is only followed by her Monday state, shifting renders her certain that it is Monday, but undecided between heads and tails. Her new evidence is neutral on *Heads & Monday* and *Tails & Monday*, so conditioning leaves these probabilities unchanged. If this seems crazy, remember that the claim is not that Beauty should trust her memories to the effect that she has not been awake after Sunday. The new probabilities are concentrated on Monday not because of Beauty's Monday evidence, but because all her previous credence went to worlds whose successors are located at Monday. (In this respect, the case of Sleeping Beauty is like Shangri La.)

Here I have assumed that Beauty's Sunday state has only one successor: her Monday state. But imagine you are in charge of the experiment. After the coin has landed tails, how do you ensure that Beauty won't be able to figure out that it is Tuesday immediately after awakening that day? It is not enough to erase her memories of Monday. You also have to erase every other trace Monday might have left. If Beauty drank on Monday, she must not have a hangover; if she broke her wrist, she must not feel pain; if she learnt Arapaho, she must have lost that ability. In effect, you have to undo everything that happened to Beauty on Monday, putting her back into the state in which she was after falling asleep on Sunday. But then her situation looks a lot like that of Fissioning Fred, whose body was scanned on Sunday and recreated from local matter on Tuesday. In Beauty's case, the recreation is made easier by the fact that the local matter – Sleeping Beauty in her post-Monday state – is already arranged not too different from the target arrangement. If Fred on Sunday has one successor on Monday and one on Tuesday, then the same should be true for Beauty.

15 This assumes that after conditioning on the information about the setup, we should give positive probability to *Heads & Tuesday* points at which the agent is asleep. It is not obvious that this is correct. The worry is even more pressing for Fissioning Fred, where the *Heads & Tuesday* possibilities don't contain an agent at all. See [Pust 2008] and [Horgan and Mahtani forthcoming] for discussion.

Suppose then that we model Sleeping Beauty as a case of epistemic fission. The update goes as follows (just like Fred's).

		H Mon	T Mon	T Tue
1/2	H Sun	1	0	0
1/2	T Sun	0	1/2	1/2
Shifting:		*1/2*	*1/4*	*1/4*
Conditioning:		*1/2*	*1/4*	*1/4*

If Beauty later finds out that it is Monday, her credence in heads rises to 2/3. This is the "Lewisian halfer" solution, from [Lewis 2001].

Unlike in the non-fission interpretation of the story, the answer this time does not depend on our anti-evidentialist account of update: we get the same solution from the evidentialist principle of inverse reflection. The reason is that in the fission scenario, Beauty knows that the predecessor of her present state is her state on Sunday, no matter what day it is and no matter how the coin has landed. But if an agent knows their predecessor's credence, shifted conditioning and the principle of inverse reflection yield the same result.[16]

It is instructive to see how this defense of Lewisian halfing fares against the host of objections that have been raised in the literature. [Piccione and Rubinstein 1997], [Dorr 2002] and [Arntzenius 2003] independently came up with the following variation of Sleeping Beauty that is supposed to undermine this solution. Suppose Beauty's memories are erased *both* on heads and on tails; however, if the coin lands heads, she gets strong evidence for *Heads & Tuesday* immediately after waking up on Tuesday. Intuitively, when she awakens on Monday, she ought to be indifferent at first between heads and tails; not finding the *Heads & Tuesday* evidence then should make her lean towards tails. So the probability of heads should be less than 1/2. This is correct:

16 On the non-fission interpretation, evidentialism combined with the principle of inverse reflection supports thirding. Let E be the total evidence Beauty receives on Monday morning. Since E is equally compatible with *Tails & Monday* and *Tails & Tuesday*, these two possibilities are presumably supported by E to roughly the same degree. Moreover, on Sunday, Beauty's credence in heads should have been 1/2, as should have been her credence in heads conditional on the assumption that she would learn E on Monday morning. Hence the shifted Sunday probability of heads conditional on E is 1/2 as well. Now *if* it is Monday, then the Sunday probability is the previous probability. By the principle of inverse reflection, it follows that $P(\textit{Heads}/\textit{Monday}) = 1/2$. Together with the fact that $P(\textit{Tails \& Monday}) = P(\textit{Tails \& Tuesday})$, this entails that $P(\textit{Heads \& Monday}) = P(\textit{Tails \& Monday}) = P(\textit{Tails \& Tuesday})$. The remaining possibility, *Heads* & *Tuesday*, is ruled out by Beauty's evidence. Hence $P(\textit{Heads}) = 1/3$.

		H Mon	H Tue	T Mon	T Tue
1/2	H Sun	1/2	1/2	0	0
1/2	T Sun	0	0	1/2	1/2
Shifting:		*1/4*	*1/4*	*1/4*	*1/4*
Conditioning:		*1/3*	*0*	*1/3*	*1/3*

But the two situations are not alike – not even with respect to Beauty's evidence after ruling out *Heads & Tuesday*. The crucial difference is that Beauty now knows that she undergoes fission on both heads and tails. In the original scenario, the probability of heads worlds in Beauty's belief space got shifted to their unique Monday successor, here it gets divided between the Monday and the Tuesday successor.

For analogous reasons, the present defense of halfing does not carry over to [Bostrom 2007]'s variation "Beauty the High Roller". On the other hand, it *does* carry over to [Titelbaum 2008]'s "Technicolor Beauty", which answers the objecting raised in [Briggs 2010]. [Elga 2000] and [Weatherson 2012] assume without further argument that Beauty's credence in heads should be 1/2 on Wednesday (Weatherson) or on Monday after learning that it is Monday (Elga). The present account gives a systematic reason why these assumptions are false.[17] Other arguments, e.g. in [Elga 2000] and [Horgan 2004], seem to consider only the evidential question, or presuppose a hybrid account of update, like [Titelbaum 2008] and [Kim 2009].

What about Dutch book arguments against Lewisian halfing, as put forward in [Arntzenius 2002], [Hitchcock 2004] and [Draper and Pust 2008]? We already know from section 6 that these arguments can't be sound: agents who follow shifted conditioning are immune to diachronic Dutch books. The alleged Dutch book goes something like this. On Sunday, Beauty is offered a bet that pays \$10 in case of tails and $\$ -9$ on heads. She will accept this bet. Whenever she awakens without remembering any post-Sunday awakening, she is offered another bet that pays \$8 in case of heads and $\$ -7$ on tails. Again, Beauty will accept. If the actual outcome is tails, the second bet gets offered twice, and the net payoff is $\$ -4$. If the outcome is heads, the bet is offered only once and the net payoff is $\$ -1$. Beauty is guaranteed to make a loss.

On the fission interpretation, the most obvious problem with this alleged Dutch book is that the two post-awakening bets are offered to different people. Whatever else one says about personal identity, it is clear that after an episode of fission, the person awakening on Monday is not identical to the person awakening on Tuesday. If Beauty's Monday successor is only interested in herself, the amount lost by the Tuesday successor therefore shouldn't be included in her total payoff.

What if we change the bets so that if either successor accepts their offer, the money is taken away from or paid out to every successor? If the two successors don't care about

17 On Wednesday, Beauty's Tails-Monday branch has come to an end, so its probability gets redistributed to the remaining branches.

one another, they should still accept their bets. And then they know that they will incur a sure loss: if the coin landed heads, there is only one person whose net loss is \$1; if it landed tails, each successor loses \$14, which yields a negative payoff even if we let both of them count the Sunday gain of \$10 against their loss. So we do have a kind of "combined Dutch book" against all the agents involved in the scenario. But the fact that rational agents are vulnerable to this kind of "Dutch book" is old news: the situation is a Prisoner Dilemma.[18]

To sum up, the answer to the Sleeping Beauty problem depends on how we interpret the question. Thirding may be the right answer to the question what evidential support Beauty's Monday evidence lends to the hypothesis of heads. On the other hand, if we ask what Beauty should actually believe, it matters whether her situation is modeled sequentially or as a case of fission. (Which of these is more plausible may depend on how exactly the scenario is spelled out.) On the sequential reading, Beauty should be certain that it is Monday, and her credence in heads should be 1/2. On the fission reading, we get the Lewisian halfer solution.

References

Carlos E. Alchourrón, Peter Gärdenfors and David Makinson [1985]: "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision". *Journal of Symbolic Logic*, (50): 510–530

Frank Arntzenius [2002]: "Reflections on Sleeping Beauty". *Analysis*, 62: 53–62

— [2003]: "Some problems for conditionalization and reflection". *Journal of Philosophy*, 100: 356–370

Nick Bostrom [2007]: "Sleeping beauty and self-location: A hybrid model". *Synthese*, 157: 59–78

Craig Boutilier [1998]: "A unified model of qualitative belief change: a dynamical systems perspective". *Artificial Intelligence*, 98: 281–316

18 Compare the following predicament. Two boxes contain either nothing or \$8 each, depending on the outcome of a fair coin toss which you did not observe. One of the boxes is given to you, the other one to somebody else. You are offered a bet that pays \$10 if the box is empty and $\$ -9$ otherwise. In addition, you have the option of opening your box. If you do so and there is money inside, you can keep it; if you open the box and it is empty, you have to pay \$7. If the other person opens her box and it is empty, you have to pay \$7 for that as well. What do you do? Answer: you should accept the bet and open the box. If your partner also opens their box, your net payoff will be either $\$ -4$ if the box is empty, or $\$ -1$ if it is not empty. You both make a sure loss.

The situation in Sleeping Beauty is actually a Twin Dilemma, and thereby a Newcomb problem. (Imagine that your partner is your twin who is very likely to do exactly what you do.) This is why the "Dutch book" could be avoided in Evidential Decision Theory, as observed in [Arntzenius 2002].

Darren Bradley [2011]: "Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty". *British Journal for the Philosophy of Science*, 62: 323–342

Rachael Briggs [2010]: "Putting a Value on Beauty". In T.S̃zabo Gendler and J. Hawthorne (Eds.) *Oxford Studies in Epistemology,* vol Vol. 3. Oxford: Oxford University Press

Cian Dorr [2002]: "Sleeping Beauty: In defence of Elga". *Analysis*, 62: 292–296

Kai Draper and Joel Pust [2008]: "Diachronic Dutch Books and Sleeping Beauty". *Synthese*, 164: 281–287

Adam Elga [2000]: "Self-locating belief and the Sleeping Beauty problem". *Analysis*, 60: 143–147

Richard Feldman and Earl Conee [1985]: "Evidentialism". *Philosophical Studies*, 48(1): 15–34

Haim Gaifman [1988]: "A Theory of Higher Order Probabilities". In B. Skyrms and W.L. Harper (Eds.) *Causation, Chance and Credence,* Dordrecht: Kluwer, 191–219

Hilary Greaves [2007]: "Probability in the Everett Interpretation". *Philosophy Compass*, 2: 109–8211

Ned Hall [2004]: "Two Mistakes about Credence and Chance". *Australasian Journal of Philosophy*, 82: 93–111

Joseph Halpern [2006]: "Sleeping Beauty reconsidered: conditioning and reflection in asynchronous systems". In Tamar Gendler and John Hawthorne (Eds.) *Oxford Studies in Epistemology, Vol.1,* Oxford University Press, 111–142

Christopher Hitchcock [2004]: "Beauty and the Bets". *Synthese*, 139: 405–420

Terry Horgan [2004]: "Sleeping Beauty Awakened: New Odds at the Dawn of the New Day". *Analysis*, 64: 10–21

Terry Horgan and Anna Mahtani [forthcoming]: "Generalized Conditionalization and the Sleeping Beauty Problem". *Erkenntnis*: –

John G. Kemeny [1955]: "Fair Bets and Inductive Probabilities". *Journal of Symbolic Logic*, 20: 263–273

Namjoong Kim [2009]: "Sleeping Beauty and Shifted Jeffrey Conditionalization". *Synthese*, 168: 295–312

Steven M. LaValle [2006]: *Planning Algorithms*. Cambridge: Cambridge University Press

David Lewis [1976]: "Survival and Identity". In Amelie O. Rorty (Hg.), *The Identities of Persons*, University of California Press, 17–40,

— [1979]: "Attitudes *De Dicto* and *De Se*". *The Philosophical Review*, 88: 513–543

— [1980]: "A Subjectivist's Guide to Objective Chance". In Richard Jeffrey (Ed.), *Studies in Inductive Logic and Probability* Vol. 2, University of California Press. Reprinted in Lewis's *Philosophical Papers*, Vol. 2, 1986.

— [2001]: "Sleeping Beauty: Reply to Elga". *Analysis*, 61: 171–176

Peter Lewis [2007a]: "Quantum Sleeping Beauty". *Analysis*, 67: 59–65

— [2007b]: "Uncertainty and Probability for Branching Selves". *Studies in History and Philosophy of Modern Physics*, 38: 1–14

Christopher Meacham [2008]: "Sleeping Beauty and the Dynamics of De Se Beliefs". *Philosophical Studies*, 138: 245–269

— [2010a]: "Two Mistakes Regarding the Principal Principle". *British Journal for the Philosophy of Science*, 61: 407–431

— [2010b]: "Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs". In Tamar Szabo Gendler and John Hawthorne (Eds.) *Oxford Studies in Epistemology, Volume 3,* Oxford University Press, 86–125

Dilip Ninan [2009]: "Persistence and the First Person". *The Philosophical Review*, 118: 425–464

Derek Parfit [1984]: *Reasons and Persons.* Oxford: Clarendon Press

John Perry [1972]: "Can the Self Divide?" *The Journal of Philosophy*, 69: 463–488

Michele Piccione and Ariel Rubinstein [1997]: "On the Interpretation of Decision Problems with Imperfect Recall". *Games and Economic Behavior*, 20: 3–24

John L. Pollock [1986]: *Contemporary Theories of Knowledge.* Towota: Rowman and Littlefield

Arthur N. Prior [1967]: *Past, Present and Future.* Oxford: Oxford University Press

Joel Pust [2008]: "Horgan on Sleeping Beauty". *Synthese*, 160: 97–101

Simon Saunders [1998]: "Time, Quantum Mechanics, and Probability". *Synthese*, 114: 373–404

— [2010]: "Chance in the Everett Interpretation". In S. Saunders, J. Barrett, A. Kent and D. Wallace (Eds.) *Many Worlds? Everett, Quantum Theory, and Reality,* Oxford: Oxford University Press

Simon Saunders and David Wallace [2008]: "Branching and Uncertainty". *British Journal for the Philosophy of Science*, 59: 293–305

Wolfgang Schwarz [2011]: "Changing Minds in a Changing World". Forthcoming in *Philosophical Studies*

Brian Skyrms [1987]: "Dynamic coherence and probability kinematics". *Philosophy of Science*, 54(1): 1–20

Paul Tappenden [2008]: "Saunders and Wallace on Everett and Lewis". *British Journal for the Philosophy of Science*, 59: 307–314

Paul Teller [1973]: "Conditionalization and observation". *Synthese*, 26(2): 218–258

Michael G. Titelbaum [2008]: "The Relevance of Self-Locating Beliefs". *The Philosophical Review*, 117: 555–606

Bas van Fraassen [1984]: "Belief and the will". *Journal of Philosophy*, 81(5): 235–256

— [1989]: *Laws and Symmetry.* Oxford: Clarendon Press

Brian Weatherson [2012]: "Ross on Sleeping Beauty". Forthcoming in *Philosophical Studies*

Alastair Wilson [2011]: "Macroscopic Ontology in Everettian Quantum Mechanics". *Philosophical Quarterly*, 61: 363–382