© 2007

Christopher J. G. Meacham

All rights reserved.

CHANCE AND THE DYNAMICS OF *DE SE* BELIEFS

BY CHRISTOPHER J. G. MEACHAM

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Frank Arntzenius

and approved by

____

____

____

____

____

New Brunswick, New Jersey

October, 2007

ABSTRACT OF THE DISSERTATION

Chance and the Dynamics of *De Se* Beliefs

by Christopher J. G. Meacham

Dissertation Director: Frank Arntzenius

How should our beliefs change over time? The standard answer to this question is the Bayesian one. But while the Bayesian account works well with respect to beliefs about the world, it breaks down when applied to self-locating or *de se* beliefs. In this work I explore ways to extend Bayesianism in order to accommodate *de se* beliefs. I begin by assessing, and ultimately rejecting, attempts to resolve these issues by appealing to Dutch books and chance-credence principles. I then propose and examine several accounts of the dynamics of *de se* beliefs. These examinations suggest that an extension of Bayesianism to *de se* beliefs will require some uncomfortable choices. I conclude by laying out the options available, and assessing the prospects of each.

Acknowledgements

While working on these topics, I have been helped, prodded, inspired and encouraged by a number of people.

My largest intellectual debt is to the members of my committee.

My advisor, Frank Arntzenius, has been extremely generous with his ideas, time and energy. I owe much of what I've learned to his creativity and eagerness to think about new things. I am very lucky to have had him as my advisor and my friend.

I owe much to Tim Maudlin, who inspired my interest these issues. Tim's penetrating insight, and his insistence on moving beyond the superficial issues, have been invaluable to both the development of this dissertation and my development as a philosopher.

I have learned much of what I know outside of the philosophy of science from John Hawthorne. His wide-ranging interests and knowledge make him a peerless resource, and his charm and humor make him a good friend.

Barry Loewer has shown me the value of getting the big picture. Barry's honest and relaxed approach to these matters has taught me much of what I know about how to do philosophy, and how to have fun doing it.

And I am grateful to Ned Hall for helpful comments and encouragement. Ned is the intellectual predecessor to most of the work on chance that appears in this dissertation.

Parts of this dissertation have been published in Meacham (2005) and Meacham (2006). For helpful comments and discussion in working out and preparing those

papers, I am indebted to Frank Arntzenius, Maya Eddon, Adam Elga, Hilary Greaves, Ned Hall, John Hawthorne, Barry Loewer, David Manley, Tim Maudlin, Adam Sennet, Ted Sider, Jonathan Weisberg and an anonymous referee.

I owe much of my education and enjoyment at Rutgers to the discussions, arguments and drinks that my peers at Rutgers have had with me. And I would like to thank my friends outside of Rutgers, who have helped to keep me sane.

I am deeply indebted to Maya Eddon. She has been a constant source of intellectual stimulation and encouragement. And she has graciously given me comments, criticisms and insights on every part of this work. Without her help, this dissertation would be much worse than it is.

And I am very grateful to Jonathan Weisberg. He has been an intellectual comrade and a beacon of philosophical sanity. I've benefited greatly from my many discussions with him. Without his friendship, I would be much worse a philosopher than I am.

Finally, I would like to thank my parents—Chuck and Jean, for instilling in me the belief that academic pursuits were worthwhile, and for too much else to much to mention; Dave and Nahid, for their support, encouragement, perspective and humor; and Sundari, who deserves at least partial parenting credit—and my siblings, Ben, Jeff, Sundari, Ratna and Shaun. I've heard that many people dislike their siblings. I can only assume that they've been less fortunate than me: I cannot imagine a group of more wonderful people.

Dedication

To the real sleeping cutie.

Table of Contents

Chapter 1
Introduction

1.1 The Problem

In standard possible worlds semantics, propositions are sets of possible worlds. To believe a proposition is to believe that your world is one of the worlds in that set. So the proposition that there are extraterrestrials is the set of worlds in which there are extraterrestrials, and to believe that there are extraterrestrials is to believe that your world is a member of that set.

A belief in a proposition is a belief about what the world is like. But in addition to beliefs about what the world is like, there are beliefs about where one is in the world. Lewis (1979) has argued that these beliefs can't be expressed in terms of possible worlds. To accommodate beliefs about where we are in the world, Lewis proposed to extend standard possible worlds semantics by introducing *centered worlds*, possible worlds paired with individuals and times.[1] A set of centered worlds is a *centered proposition.*[2] To believe a centered proposition is to believe that your current centered world is one of the centered worlds in that set. So the centered proposition that it's 9 am is the set of centered worlds at which it's 9 am, and to believe that it's 9 am is to believe that your current centered world is a member of that set.

Following Lewis, call beliefs that can be expressed in terms of possible worlds

[1] This is one way to make sense of centered worlds, anyway. For a more detailed discussion of some of these issues, see chapter 4.

[2] Lewis (1979) himself calls them *properties.*

de dicto beliefs, and beliefs that can be expressed in terms of centered worlds *de se beliefs*. In his 1979 paper Lewis raises the question of what happens to Bayesian decision theory when we consider *de se* beliefs instead of *de dicto* beliefs. His answer is a natural one:

> "Very little. We replace the space of worlds by the space of centered worlds, or by the space of all inhabitants of worlds. All else is just as before."[3]

However, this answer is untenable. When you update your beliefs using standard Bayesian conditionalization your certainties are permanent: if you're certain a proposition is true before updating then you'll be certain it's true after updating. So on the account Lewis suggests, if you're certain that a centered proposition is true you will always remain certain that it's true. But suppose you're looking at a clock you know is accurate. If the clock reads 9 am, then you're certain of the centered proposition that it's 9 am. Given Lewis's suggestion, since you're certain of the centered proposition that it's 9 am when the clock reads 9 am, you should always remain certain that it's 9 am. So you should remain certain that it's 9 am a minute later, when the clock reads 9:01 am. Obviously, this is not how our beliefs should be updated.[4] We need a more sophisticated dynamics for *de se* beliefs.

I've followed Lewis in using centered worlds to model self-locating beliefs, but nothing hangs on this. The problem persists regardless of how we choose to model the beliefs in question. Consider the belief that W is a precise description of what the world is like, and that you are individual i at location l at time t. If a method M of representing the objects of belief cannot capture such beliefs, then M is too coarse grained to capture the kinds of beliefs we want to consider. On the other hand, if a method M of representing the objects of belief can capture

[3] Lewis (1979), p. 149.

[4] Arntzenius (2003*a*), Halpern (2005) and Hitchcock (2004) have noted this problem with extending standard conditionalization to *de se* beliefs.

such beliefs, then we can take centered worlds to correspond to these M-beliefs, and understand discussion about what our beliefs in centered worlds should be as discussion about what our beliefs in these M-surrogates should be, instead.

This work is an attempt to figure out the correct dynamics for *de se* beliefs. In the chapters that follow I will look at several accounts of the dynamics of *de se* beliefs, and assess their virtues and shortcomings. Before I look at these accounts, however, some preliminary work needs to be done. In the rest of this chapter I will lay out some of the necessary background. In the next section I will sketch the standard Bayesian account, in both its classical and modern formulations. In the third section I will describe my views on the methodology underlying the Bayesian project, and describe how this work fits into that project. In the fourth section I will describe a paradigmatic case of self-locating belief, the sleeping beauty case, which will serve as a useful example throughout our discussion. In the fifth section I will provide a more detailed sketch of how the rest of this work will proceed.

1.2 Bayesianism

We can divide the Bayesian theory into two parts: assumptions about the agents to which the theory applies, and a normative claim about how such agents ought to update their beliefs upon receiving evidence.

The agents to which Bayesianism applies satisfy the following conditions:

A1. The agent's epistemic state at a time can be represented by a probability function over a space of possibilities. These values, called *credences* or *degrees of belief*, indicate the subject's confidence that the possibility is true, where greater values indicate greater confidence.[5]

[5] So 0 indicates virtual certainty that a possibility is false, 1 indicates virtual certainty that a possibility is true. Why *virtual* certainty instead of certainty? Because it's not clear we want

A2. The agent's evidential state at a time can be represented by a set of possibilities. The possibilities in the set are the possibilities compatible with the evidence the agent has.[6]

A3. The space of possibilities in question is the space of possible ways the world could be (in possible worlds jargon, the space of possible worlds).

The normative part of the Bayesian theory claims that agents who satisfy A1-A3 ought to have credences which satisfy a particular constraint. Let cr_E stand for the credence function of a subject with evidence E. (I'll sometimes omit the subscript when the content of the subject's evidence is irrelevant.) The classical formulation of the Bayesian constraint is:

Bayesianism (Classical): If a condition-satisfying agent with credences cr gets evidence E, then her new credence function cr_E should be:

$$cr_E(\cdot) = cr(\cdot|E), \tag{1.1}$$

where the usual definition for conditional probability is being employed.[7]

It's often convenient to formulate the Bayesian constraint in another way. Here's the idea. If you're a good Bayesian, the classical Bayesian formula gives you the relation between your current credences and your previous credences, the

a credence of 1 to indicate certainty. After all, there's a chance of 1 that an infinite sequence of coin tosses won't all land heads, and it seems our credences should line up with the chances. But we shouldn't be *certain* that the coins won't all land heads, since it's possible they could. (I discuss this issue further in chapter 4.)

[6] Two comments. First, depending on one's account of evidence, this condition might be unnecessary. For example, this condition isn't necessary if, like Howson and Urbach (1993), one takes one's evidence to just be the possibilities to which one hasn't assigned a credence of 0. (Note that this proposal weakens the normative impact of Bayesianism. Bayesianism is a stronger normative constraint if we adopt a substantive account of evidence, such as your evidence being the set of possibilities compatible with your subjective state.)

Second, it's unclear whether this way of treating evidence can accommodate vagueness. It depends, though, on one's treatment of vagueness, as well as one's account of evidence. (This issue is discussed in chapter 4.)

[7] The conditional probability of A given B is $p(A|B) = \frac{p(AB)}{p(B)}$.

credences you had before you got your last piece of new evidence. Likewise, it gives you the relation between your previous credences and the credences you had before that. And if we push back far enough, to earlier and earlier credence functions, we can imagine ending up with some hypothetical initial credence function: your credences before you had any evidence at all. The modern formulation of Bayesianism works directly with these hypothetical initial credence functions, or *hypothetical priors*.

Bayesianism (Modern):

(i) A condition-satisfying agent with evidence E and hypothetical priors hp should have the following credences:

$$cr_E(\cdot) = hp(\cdot|E). \tag{1.2}$$

(ii) A condition-satisfying agent should have permanent certainties. I.e., if her credence in a proposition becomes 0 or 1, it should remain 0 or 1. (Given (i), not losing information is a necessary and sufficient condition for (ii). So given (i), we can replace (ii) with the following clause: a condition-satisfying agent with old evidence E and new evidence E' will always be such that $E' \Rightarrow E$.)

Note that the classical formulation doesn't require the second clause, because (ii) is entailed by the classical Bayesian formula.[8]

One advantage of the modern formulation of Bayesianism is that it provides an easy way to characterize the difference between objective and subjective Bayesians. Objective Bayesians hold that there is only one rationally permissible set of hypothetical priors. Extreme subjective Bayesians hold that any set of

[8] If $cr(A) = 0$, then $cr_F(A) = \frac{cr(AF)}{cr(F)} = 0$, regardless of what new evidence F one receives.

(probabilistically coherent) hypothetical priors is rationally permissible. Moderate subjective Bayesians hold something between these two views: more than one set of hypothetical priors is rationally permissible, but not all of them are. Other advantages come from the ease with which it allows one to discuss various issues, such as the problem of old evidence (see Earman (1992) and Howson and Urbach (1993)), inductive frameworks (see Strevens (2004)), confirmation relations (see Maher (2006)), the chance-credence relation (see Lewis (1986*b*) and Hall (1994)), and the Doomsday argument (see Bartha and Hitchcock (N.d.)).

That said, the modern formulation of Bayesianism faces some worries as well. In particular, it requires an account of hypothetical priors. Here are three ways to make sense of them.

First, we might take hypothetical priors to be the initial credence function of a condition-satisfying agent before she has any evidence. On this view, calling them "hypothetical" is a misnomer, since they're not hypothetical—they're actual. This gives us a concrete account of what hypothetical priors are. But it's not clear that agents like us were ever in a state without evidence. If we were never in such a state, then none of us have hypothetical priors. So if we adopt this approach, it's natural to add the following condition to A1-A3: "The agent's initial credences were not informed by any evidence". The extent to which this addition is problematic depends on how one understands the role of idealizations in the Bayesian project. I'll present my views on the matter in the next section.

Second, we might take hypothetical priors to be a "normative stamp" on each condition-satisfying agent. That is, an agent's hypothetical priors encode certain normative facts which place normative constraints on her beliefs. This view is natural for objective Bayesians, who will hold that all condition-satisfying agents have the same normative stamp. But this view can be held by subjective

Bayesians as well: a condition-satisfying agent's credences are bound by her hypothetical priors, but different subjects can be bound in different ways. In any case, there are further details to be filled in about the kinds of facts that determine the values of an agent's hypothetical priors, what priors are had by which agents, and so on.

Third, we can adopt a deflationary account of hypothetical priors by understanding (1.2) as a kind of functional constraint. On this account, the right way to understand (1.2) is this: an agent's credences and evidence at every time should be such that there exists a probability function "*hp*" which yields each of those credences when conditioned on the corresponding evidence. In many cases "*hp*" won't even be unique—there will be several such functions which yield the credences of the agent in the appropriate way.[9] This view is natural for extreme subjective Bayesians, since on this view (1.2) becomes nothing but a constraint on how our credences at different times ought to be correlated. But this view can be held by objective Bayesians and moderate subjective Bayesians as well: they can maintain a deflationary understanding of hypothetical priors by adding in the constraints on priors as part of the functional constraint (1.2) imposes. I.e., a moderate subjective Bayesian who holds that hypothetical priors should satisfy constraint X will understand (1.2) this way: an agent's credences and evidence at every time should be such that there exists a probability function "*hp*" satisfying conditions X which yields each of those credences when conditioned on the corresponding evidence.

I won't choose between these alternatives here. By and large, all of these understandings are compatible with the discussion to follow. In places where the account of hypothetical priors one adopts makes a difference, I will note it.

[9] In particular, if an agent has a 0 credence in some possibilities at every time, then multiple *hp* functions can be used to generate her credences.

1.3 Methodology

Bayesianism applies to condition-satisfying agents. But condition-satisfying agents are highly idealized. A1 requires agents to have *precise* credences in *all* possibilities, and requires these credences to satisfy the probability axioms. It's unlikely that any of these criteria are met by agents like us, who appear to have vague credences in some possibilities, no credences at all in others, and whose credences often appear to violate the probability axioms. A2 assumes that an agent's evidence can be adequately represented by a set of possibilities. But one might think that evidence needs to be represented by something more sophisticated. (Jeffrey conditionalization, for example, requires a more complicated picture of evidence.[10]) A3 assumes that we can take the possibilities that form the objects of belief to be something like ways the world could be. But as we've seen, this cannot accommodate self-locating beliefs. Likewise, this cannot accommodate beliefs in claims that are necessarily true, such as the claim that Hespherous is Phosphorous, that water is H_2O, that $\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B)$, and so on.

The fact that the condition-satisfying agents are highly idealized is often raised as a complaint against Bayesianism: "Bayesianism makes a number of idealizations about agents which are false for people like us. But we're interested in finding out what *we* ought to believe, and Bayesianism doesn't tell us anything about that. So why should we care about it?"

I don't find this criticism compelling. I take the methodology underlying the Bayesian project to be similar to the methodology used in the sciences. In the end, we want a theory that is perfectly precise and doesn't make any idealizations. But trying to come up with a perfectly precise theory right away is an intractable project. The space of potential theories is too big to explore without further guidance. This is where idealization come in. It is tractable to figure out what

[10] See Jeffrey (1983).

the correct theory is in highly idealized contexts. And by looking at how to extend idealized theories to less idealized contexts, these theories can guide our exploration of the space of potential theories.

So I agree that, in the end, we'd like an account which can make sense of degrees of belief which aren't precise, agents who aren't modally and logically omniscient, and so on. But we can't get there all at once. We need to start with something tractable, figure out what to say there, and then see how to extend the theory as more and more of these idealizations are relaxed.

A considerable amount of work has already been done on trying to relax the standard Bayesian idealizations. There have been proposals that replace precise credences with vague ones, proposals that allow agents to lack credences in some possibilities, proposals that allow credences to violate the probability axioms, proposals which modify the picture of evidence, and so on. This work is another attempt to relax a standard Bayesian idealization: it explores the consequences of relaxing A3, and extending Bayesianism to self-locating beliefs as well as beliefs about what the world is like.

1.4 Sleeping Beauty

A paradigmatic case of *de se* belief change, which will serve as a useful example throughout our discussion, is the sleeping beauty case:

> *The Sleeping Beauty Case:* Some researchers are going to put Beauty to sleep on Sunday night, and then flip a coin. If heads comes up they will wake her up on Monday morning. If tails comes up they will wake her up on Monday morning and Tuesday morning. And in between Monday and Tuesday, while she's sleeping, they will erase the memories of her waking.
>
> What should Beauty's credences be when she wakes up on Monday morning? And what should Beauty's credences become if she's told that it's Monday?

In the work that follows, we'll consider several accounts of the dynamics of *de se* beliefs. Although all these accounts differ in some respects, they can be naturally divided in accordance with how they assign credences to heads and tails in the sleeping beauty case.

The first group of accounts were constructed by Halpern (2005) and Meacham (2006) to yield the response given by Elga (2000) to the sleeping beauty case. These accounts assign $\frac{1}{3}/\frac{2}{3}$ to heads/tails when Beauty wakes up, and $\frac{1}{2}/\frac{1}{2}$ after she is told it's Monday. The second group of accounts assigns $\frac{1}{2}/\frac{1}{2}$ to heads/tails when Beauty wakes up, and $\frac{1}{2}/\frac{1}{2}$ after she is told it's Monday. This group includes views defended in Halpern (2005) and Meacham (2006), as well as *temporal successor conditionalization* (chapter 7) and a *combination account* (chapter 8). The third group assigns $\frac{1}{2}/\frac{1}{2}$ to heads/tails when Beauty wakes up, and $\frac{2}{3}/\frac{1}{3}$ after she is told it's Monday. This group consists of a view constructed in Meacham (2006) to yield the response offered by Lewis (2001) to the sleeping beauty case, and *epistemic successor conditionalization* (chapter 7).

1.5 Outline

The rest of this work will proceed as follows. In chapters 2 and 3 I'll look at some attempts to resolve the sleeping beauty case directly, without settling on a detailed account of the dynamics of *de se* beliefs. In chapter 2 I'll assess the prospects for resolving the sleeping beauty case by appealing to a chance-credence principle such as the Principal Principle. In the process of getting straight on the form and status of the correct chance-credence principle, I'll also examine some broader issues regarding a theory of chance. In chapter 3 I'll examine some attempts to settle the sleeping beauty case using Dutch book arguments.

The remainder of this work will focus on specific accounts of the dynamics of *de se* beliefs. In chapter 4 I'll lay out a formal framework for discussing features of *de se* beliefs that will be employed in the discussion that follows. Then I'll look at the motivations for, and prospects of, several accounts. In chapter 5 I'll examine three accounts of the dynamics of *de se* beliefs that employ hypothetical priors, and the problems they encounter. Motivated by these problems, in chapter 6 I'll consider some principles that place constraints on the credence assignments of rational updating rules. In chapter 7 I'll look at two accounts which satisfy some of these principles. I conclude in chapter 8, where I'll discuss the role of updating rules as guides for epistemic subjects, and the tension between providing guidance and satisfying other intuitively desirable features. Then I'll offer an assessment of the options available to us in light of the prior discussion.

A note on style before we proceed. In a number of places I've had to choose whether to include technical material in the text, or to place it in an appendix. Moving such material to the appendix makes the text more readable, but in some cases the technical results are the bulk of the content of a section. I've decided to err in favor of readability. So I've consistently moved the technical parts of the discussion into appendices.

Chapter 2
Chance

2.1 Introduction

In the sleeping beauty case it's uncontentious that your credence should line up with the chance of heads/tails on Sunday. But should your credences still line up with these chances when you wake up on Monday? Lewis (2001) and Dorr (2002) note that the answer to this question seems to hang on how credence and chance are related. After all, if we were given an account of how they ought to be related, it seems we could figure out what the answer to the sleeping beauty case should be. If the account told us that our credence should still line up with the chance of heads/tails on Monday morning, for example, then the $\frac{1}{2}/\frac{1}{2}$ response would be correct.[1] So if we want to resolve the sleeping beauty case, it seems we should first get clear on the principle relating credence and chance.

But this won't be our only concern. In what follows, I'll look at the broader issue of what an adequate theory of chance should look like. This includes, but is not exhausted by, the question of what the correct chance-credence principle is.

[1] At this point, some have had the following thought: "The chances will be 0 and 1 on Monday morning, since the coin has landed. And no one thinks our credences should be 0 or 1. So it's not clear that the chance-credence principle directly applies to the sleeping beauty case, as it is usually formulated." This worry needn't concern us, for two reasons. First, the case is easily changed to accommodate this worry: we can have the coin toss occur on Monday night instead of Sunday night. Then the current chance of heads/tails will still be $\frac{1}{2}/\frac{1}{2}$, and the case proceeds as before. The second and deeper reason this shouldn't concern us is that this thought relies on a mistaken understanding of the Principal Principle. The Principal Principle tells you to align your credences with the chances at a time that you believe obtain, and which you don't have inadmissible evidence with respect to. It doesn't say anything about how that time relates to what time it is now: the time of evaluation is irrelevant. (Peter Vranas makes a mistake of this kind in his formulation of the Principal Principle; see section 2.2.2.)

I do this in part because these issues are bound together in various ways, and so are easier to assess in unison.

A comprehensive theory of chance should address a number of questions. Dividing them by topic, I take the central questions to be these:

1. A metaphysical account of chance
    (a) What is the nature of chance? (Are they frequencies? Propensities? Or what?)
    (b) What is the structure of chance?
    (c) What is the structure of chance theories?
2. An account of the relation between credence and chance
    (a) What is the correct chance-credence principle?
    (b) What is the status of this chance-credence principle? (Does it provide an analysis of chance? Does it contain everything we know about chance?)

These divisions are rough and ready; obviously these issues overlap.

In what follows I'll take a critical look at how the canonical theory of chance, proposed by Lewis (1986*b*), addresses these questions. I'll then propose an alternative theory of chance which remedies some of the defects of Lewis's account. However, question 1A—what is the nature of chance?—will be largely absent from my discussion. This is at odds with much of the recent literature. Much of the discussion of Lewis's theory has focused on whether his Humean answer to this question is tenable. The work done in this chapter is largely orthogonal to this issue, so I've tried my best to avoid it. Unfortunately, the issue of Humeanism so pervades the literature on chance that it is impossible to avoid it completely. In this paper I attempt the following compromise: in the body of the chapter I will try to sidestep issues regarding Humeanism, and I leave an assessment of the (lack of) implications my discussion has on Humeanism to Appendix 9.1.

The rest of this chapter will proceed as follows. In section 2.2 I'll sketch Lewis's metaphysical theory of chance and his account of the relation between credence and chance. In section 2.3 I'll evaluate Lewis's account of the relation between credence and chance. In section 2.4 I'll propose an alternative account of this relation. In section 2.5 I'll evaluate Lewis's metaphysical account of chance. In section 2.6 I'll propose an alternative metaphysical account. In section 2.7 I'll return to the question with which we started, and assess what the correct chance-credence principle tells us about the sleeping beauty problem.

2.2 Lewis's Theory of Chance

A theory of chance has two parts: a metaphysical account of chance, and account of the relation between chance and credence. Let's look at each of these in turn.

2.2.1 Lewis's Metaphysical Account of Chance

In the introduction I raised three questions that a comprehensive metaphysical account of chance should answer. The third question—what is the structure of chance theories?—is something Lewis says very little about. The first question—what is the nature of chance?—is something Lewis says much about, but it is orthogonal to the issues we're interested in. This leaves us with the second question—what is the structure of chance?

Lewis makes a number of substantial claims about the structure of chance. Since we'll be looking at these claims in detail, it will be useful to present and label each claim separately.

L1. Every possible chance assignment can be encoded by a single function, *ch*, which takes a *grounding argument* G and spits out a probability function

$ch_G(\cdot)$.

L2. Chances are assigned to ways the world could be; i.e., *de dicto* propositions. So the chance distributions $ch_G(\cdot)$ are probability functions over *de dicto* propositions.[2]

I'll say that a chance distribution $p(\cdot)$ *obtains at world* w iff some G obtains at w such that $ch_G(\cdot) = p(\cdot)$. $\langle ch(\cdot) = p(\cdot)\rangle$ is the proposition that $p(\cdot)$ obtains at this world.

L3. The chance distributions $ch_G(\cdot)$ assign values to every proposition (or at least every proposition to which an idealized credence function assigns values).

L4. The grounding argument (or *grounds*) of the chance distribution is a conjunction of a complete chance theory T and a complete history up to a time at a world where that chance theory holds, H. Note that:

(i) T and H entail the chance distribution they ground.[3]

(ii) The chance distributions supervene on T and H.[4]

(iii) Since T and H can be uniquely picked out by a time and a world, we can also take ch to be a function that takes time and world pairs t, w and spits out a chance distribution.

[2] This is given as a simplifying assumption by Lewis (1986*b*). But in later work, such as Lewis (2004), he retains this as a substantive assumption.

[3] Suppose $ch_{TH}(\cdot) = p(\cdot)$. It follows that if TH is true (obtains at this world) then $\langle ch(\cdot) = p(\cdot)\rangle$ is true.

[4] Proof: Suppose otherwise. Then there exists two worlds, w_1 and w_2, such that (a) $\forall i(w_1 \Rightarrow T_iH_i)$ iff $(w_2 \Rightarrow T_iH_i)$, and such that (b) $((w_1 \Rightarrow \langle ch(\cdot) = p(\cdot)\rangle)$ and $(w_2 \not\Rightarrow \langle ch(\cdot) = p(\cdot)\rangle))$. But if there is no T_jH_j at w_1 such that $T_jH_j \Rightarrow \langle ch(\cdot) = p(\cdot)\rangle$, then $\langle ch(\cdot) = p(\cdot)\rangle$ can't obtain at w_1, since $\langle ch(A) = x\rangle \Leftrightarrow \bigvee_{(i|ch_{T_iH_i}(A)=x)} T_iH_i$, so (b) is false. And if there is a T_jH_j at w_1 such that $T_jH_j \Rightarrow \langle ch(\cdot) = p(\cdot)\rangle$, then $w_2 \Rightarrow \langle ch(\cdot) = p(\cdot)\rangle$ too, since T_jH_j holds at w_2, so (b) is also false. By *reductio*, the chance distribution facts supervene on the TH facts.

I'll say that a chance distribution $p(\cdot)$ *obtains at world w at time t* iff some TH obtains at w such that H is a history up to t and $ch_{TH}(\cdot) = p(\cdot)$. $\langle ch_t(\cdot) = p(\cdot)\rangle$ is the proposition that $p(\cdot)$ obtains at this world at t.

L5. If $H \Rightarrow A$ then $ch_{TH}(A) = 1$, if $H \Rightarrow \neg A$ then $ch_{TH}(A) = 0$.[5] ("The past is no longer chancy.")

L6. For any chance theory T and history H that obtain at a world where determinism holds, $ch_{TH}(A)$ is 0 or 1.[6] ("Determinism and chance are incompatible.")

Considered in isolation, L5 and L6 may seem somewhat *ad hoc.* But Lewis has principled reasons for adopting them. Given L1-L4, Lewis's original chance-credence principle and some additional assumptions, one can derive L5 and L6 (see Appendix 9.2 for details). The additional assumptions needed are not entirely innocuous (Lewis later rejected his original version of the chance-credence principle, for example, due to purported incompatibilities with Humeanism[7]), so I've included L5 and L6 as separate assumptions.

L3 also deserves some further comment. Lewis (1986*b*) assesses the claim that chance distributions assign values to every proposition. Although he is clearly sympathetic to this claim, he holds back from endorsing it: "It is only caution, not any definite reason to think otherwise, that stops me from assuming that chance of truth applies to any proposition whatever".[8] But if he explicitly declines to endorse the claim that chance distributions assign values to every proposition, why am I adding L3 to the claims Lewis makes about chance?

[5] Alternatively: If $H_{tw} \Rightarrow A$ then $ch_{tw}(A) = 1$, if $H_{tw} \Rightarrow \neg A$ then $ch_{tw}(A) = 0$.

[6] Alternatively: If w is a deterministic, then for all t and A, $ch_{tw}(A)$ is 0 or 1.

[7] See Lewis (1994).

[8] Lewis (1986*b*), p.91.

The key is the parenthetical addition to L3. Although Lewis does not claim that chance distributions assign values to every proposition, he does endorse L3. In the postscript to this paper, Lewis spells out the source of his caution in the above statement:

> "My reason for caution was not that I had in mind some interesting class of special propositions—as it might be, about free choices—that would somehow fail to have well-defined chances. Rather, I thought it might lead to mathematical difficulties to assume that a probability measure is defined on all propositions without exception. In the usual setting for probability theory—values in standard reals, sigma-additivity—that assumption is indeed unsafe... I did not know whether there would be any parallel difficulty in the nonstandard setting [which Lewis endorses]... Plainly this reason for caution is no reason at all to think that the domains of chance distributions will be notably sparser than the domains of idealized credence functions."[9]

So with the parenthetical addition, L3 is faithful to Lewis's claims about chance.

2.2.2 Lewis's Account of the Relation Between Credence and Chance

Now let's look at Lewis's account of the relation between credence and chance.

Lewis's Chance-Credence Principle

Lewis famously proposed that the correct relation between credence and chance is provided by the *Principal Principle*:

$$\text{PP}_1 : hp(A|\langle ch_t(A) = x\rangle E) = x, \text{ if } \langle ch_t(A) = x\rangle E \text{ is admissible at } t. \quad (2.1)$$

[9] Lewis (1986*b*), p.132.

(On Lewis's (1986) original formulation, the admissibility clause was "if E is admissible relative to t". The admissibility clause given here is the one described in Lewis (1994).[10])

What is *admissible* evidence? Lewis never provided a precise account of when evidence is admissible, but the intuitive idea is that evidence is admissible if it is not relevant to the outcome of the chance event in question. As a rule of thumb, he takes information about the past to be admissible, and information about the future to not be admissible. But as Lewis notes, these rules have exceptions. Crystal balls or oracles may make past evidence inadmissible, and future evidence irrelevant to the outcome of the chance event may be admissible.[11]

Lewis (1986*b*) assumes that knowledge of a chance or its grounds is admissible. relative to the time of that chance assignment. This assumption is vital to the work he provides there: many of the arguments and every derivation that he presents requires this assumption to go through. Indeed, without this assumption the Principal Principle is vacuous, since the admissibility clause will never be satisfied. So I'll follow Lewis in making this assumption throughout the discussion.

It should be noted, however, that this assumption is not entirely innocuous. Lewis came to believe that this assumption was incompatible with Humean accounts of chance, and suggested replacing the Principal Principle with a variant, the "New Principle". I'll largely ignore these issues, for two reasons. First, Vranas (2002) has shown that this assumption *is* compatible with Humeanism after all, so Lewis need not have rejected the Principal Principle (see Appendix 9.1). Second, the New Principle is the same as the Principal Principle in the ways relevant to our discussion.

PP_1 is one way to formulate of the Principal Principle. But we can also

[10] See Lewis (1994), p. 238.

[11] Or at least "very nearly admissible"; see Lewis (1994), p.242.

formulate the principle in terms of the grounding argument TH directly:

$$\text{PP}_2 : hp(A|TH) = ch_{TH}(A) \tag{2.2}$$

Given the assumptions Lewis (1986*b*) makes, PP$_1$ and PP$_2$ are equivalent. In particular, given L1-L4 and some assumptions about admissibility, PP$_1$ can be derived from PP$_2$, and vice versa (see Appendix 9.3). (In fact, Lewis notes that these derivations only go through for cases where $\langle ch_t(A) = x\rangle$ is equivalent to a finite disjunction of TH terms.[12] But he states that this lapse is unlikely to be important, and I am inclined to agree. Following his lead, I will usually restrict my attention to finite cases from now on.)

I've provided what I take to be the most perspicuous formulations of Lewis's Principal Principle. But Lewis's presentation of these principles varies somewhat, as does the way they are presented in literature, so there is room for disagreement. For example, one might quibble about how I've formulated the admissibility clause of PP$_1$. (Should it just be E that needs to be admissible, or everything in the consequent of the conditional? Should E be assessed relative to a time as Lewis originally suggested, relative to a time and a proposition as Thau (1994) has suggested, or relative to something else?)

I think most of these differences are reasonable, harmless or both. But some of these disagreements are not harmless. Some of the formulations of the Principal Principle that have appeared in the literature are inconsistent, or fail to apply in important cases. Let's go over two mistakes in formulating the Principal Principle that have appeared in the literature, and look at why they're problematic.

[12] See Lewis (1986*b*), p.100.

Mistake 1: Conditional Credence

The intuitive motivation behind the Principal Principle is something like this: if a subject believes that an event has a given chance of occurring, this should have some bearing on her credence that the event will occur. Neither of the formulations of the Principal Principle given above provide a direct constraint of this kind. These principles place constraints on hypothetical priors (or "reasonable initial credence functions", as Lewis called them), not credences. So are these principles giving us what we want?

They are if, following Lewis, we assume something like Bayesianism is true. A modern formulation of Bayesianism relates our credences to our hypothetical priors:

$$cr_E(A) = hp(A|E) \tag{2.3}$$

With this in hand, each of the two versions of the Principal Principle can be turned into a constraint on our credences:

$$cr_{\langle ch_t(A)=x\rangle E}(A) = x, \text{ if } \langle ch_t(A) = x\rangle E \text{ is admissible at } t. \tag{2.4}$$

$$cr_{TH}(A) = ch_{TH}(A) \tag{2.5}$$

Now, if one wants to formulate the chance-credence constraint directly in terms of credences, one might adopt these rules in place of Lewis's original formulations of the Principal Principle. But we must make sure that the substitutions we employ are those just given. We'll run into trouble if, like Strevens (1995) and Vranas (2002), Vranas (2004), we simply replace the conditional priors in Lewis's formulation with conditional credences:

$$cr(A|\langle ch_t(A) = x\rangle E) = x, \text{ if } \langle ch_t(A) = x\rangle E \text{ is admissible at } t. \tag{2.6}$$

$$cr(A|TH) = ch_{TH}(A) \tag{2.7}$$

Both of these principles lead to inconsistencies. To see this, pick a T, H and A such that $ch_{TH}(A) = x \neq 1$, and consider a subject whose total evidence is THA. Then for (2.6):

$$1 = cr_{THA\langle ch_t(A)=x\rangle}(A) = x \neq 1. \tag{2.8}$$

Likewise, for (2.7):

$$1 = \frac{cr_{THA}(THA)}{cr_{THA}(TH)} = cr_{THA}(A|TH) = ch_{TH}(A) = x \neq 1. \tag{2.9}$$

So we should avoid attempting to formulate the Principal Principle in terms of conditional credence.

Mistake 2: Time-Indexing

Consider a subject who knows the chance at several different times, t_1-t_n, of some outcome A. And suppose that the chance of this outcome is different at each of these times. Which of these chances should her credences in the outcome line up with? And how does the Principal Principle deliver this result?

Here is how Lewis would answer the question. We only set our credence in accordance with a chance if we have no inadmissible evidence with respect to that chance. If we know the chance of A at several times, the knowledge of what the chance will be at future times will be inadmissible with respect to the chances that held at earlier times. Thus the first formulation of the Principal Principle will only apply to the latest chance of A that we know of. So assuming we don't have any other evidence that's inadmissible, our credence in A should line up with the chance at t_n.

Note that this answer to the question is independent of the time at which the agent is located. The time at which the agent is considering which credences to adopt might be before t_1, after t_n, or some time in-between. But this has no bearing on what his credences should be in the case above. (For those familiar with the literature, this is essentially the point Lewis (1986*b*) makes in his response to Henry Kyburg in the postscript to this paper.)

Here is a different answer to the question. We want our *current* credence to line up with what we think the *current* chance is. So we should add a time index to the credence function on the left hand side of the Principal Principle, and (2.4) should really be:

$$cr^t_{\langle ch_t(A)=x\rangle E}(A) = x, \text{ if } \langle ch_t(A) = x\rangle E \text{ is admissible at } t. \tag{2.10}$$

This is essentially what Vranas (2002) and Vranas (2004) presents as Lewis's Principal Principle. ('Essentially' because he also formulates the principle in terms of conditional credence, as described above.) But this isn't what Lewis wanted, with good reason. Vranas's Principal Principle won't allow us to make all of the inferences we'd like to make.

Say you're told that a fair coin was flipped a hundred years ago, before you were born. You don't know anything else about the coin toss, however, or how it came out. What should your credence be that the coin came up heads?

It seems your credence should be $\frac{1}{2}$. And this is what Lewis's Principal Principle tells us to believe: you know the chance of the coin coming up heads was $\frac{1}{2}$, and you don't have any evidence that's inadmissible relative to that chance. But on Vranas's version of the Principal Principle, this won't follow. You don't know the *current* chance of the coin coming up heads—you don't know whether it's 0 or 1—and so Vranas's Principal Principle won't apply.

(But if your credences in heads and tails were appropriately constrained at the time of the coin toss, and you update using the Bayesian rule, won't you end up with the right credences later on? If so, doesn't this get Vranas out of the above problem? Yes, you will end up with the right credences, but no, this won't get Vranas out the problem. In the case I describe, the coin is flipped before you were born. So your credences in heads and tails can't have been appropriate constrained at the time of the coin toss. And once that time has passed, it's too late.

It's natural at this point to think that a shift to a constraint on hypothetical priors is called for. If we formulate the chance-credence principle as a hypothetical prior constraint, we don't need to worry about when you were born. But this move isn't available to Vranas. Because the credences in his chance-credence principle are time indexed, there's no obvious way to formulate an equivalent principle in terms of hypothetical priors.)

The Status of the Chance-Credence Principle

Lewis evaluates two interesting claims about the status of the Principal Principle. First, he states that the Principal Principle encodes everything we know about chance. This is at least *prima facie* surprising, since several of the claims Lewis makes, like L1-L6, don't seem to follow from the Principal Principle. Second, Lewis assesses the prospects of using the Principal Principle to provide an analysis of chance. If one thinks that the Principal Principle tells us everything we know about chance, then attempting to use the Principal Principle to provide an analysis of chance is a natural next step. After all, why think there's more to chance than we know about it? I.e., why think there is anything more to being a chance than playing the appropriate role in the Principal Principle? Here, though, Lewis is more cautious. He has some suggestions for how one may try to carry out such an analysis, but he concedes that the project looks like a difficult one. (We'll look at his comments on the prospects for such an analysis in the next section.)

2.3 Evaluating Lewis's Chance-Credence Principle

With Lewis's theory of chance in hand, we can now turn to evaluating its components. We'll start with the cornerstone of his account, the Principal Principle.

2.3.1 The Principal Principle and Time

Lewis (1986*b*) formulates PP$_1$ using time-indexed chances:

$$\text{PP}_1 : hp(A|\langle ch_t(A) = x\rangle E) = x, \text{ if } \langle ch_t(A) = x\rangle E \text{ is admissible at } t. \tag{2.11}$$

If we drop the time index, we get the following principle, which I'll call PP$_3$:

$$\text{PP}_3 : hp(A|\langle ch(A) = x\rangle E) = x, \text{ if } \langle ch(A) = x\rangle E \text{ is admissible.}^{13} \tag{2.12}$$

$\langle ch(A) = x\rangle$ is the *de dicto* proposition containing every world at which there is a history H and the chance theory T such that $ch_{TH}(A) = x$.

How are these two principles related? PP$_1$ is related to PP$_3$ in the same was as PP$_2$ is related to PP$_1$. The difference between PP$_2$ and PP$_1$ is that PP$_2$ is formulated in terms of particular grounds for chance distributions, TH, while PP$_1$ is formulated in terms of $\langle ch_t(A) = x\rangle$, a large disjunction of TH terms. The difference between PP$_1$ and PP$_3$ is that PP$_1$ is formulated in terms of $\langle ch_t(A) = x\rangle$, while PP$_3$ is formulated in terms of $\langle ch(A) = x\rangle$, a large disjunction of $\langle ch_t(A) = x\rangle$ terms. And the assumptions which allows PP$_1$ to be derived from PP$_2$ and vice versa, also allow PP$_3$ to be derived from PP$_1$ and PP$_2$, and vice versa (see Appendix 9.4 for details).

So PP$_3$ gives us a third formulation of the Principal Principle, equivalent to the first two but without time indexed chances. The appearance of the time-index in PP$_1$ is, in fact, superfluous.

[13] Where this kind of admissibility is defined as follows: $\langle ch(A) = x\rangle E$ is admissible iff (i) $\langle ch(A) = x\rangle E \neq \emptyset$, and (ii) $\forall t$ either (a) $\langle ch_t(A) = x\rangle E$ is admissible relative to t, or (b) $\langle ch_t(A) = x\rangle E = \emptyset$.

This may seem surprising at first. Here's a natural gut-reaction response:

> That can't be right. Consider an agent who knows nothing except a single fact about chance—that at some time the chance of A will be x. PP$_3$ says that this agent's credence in A ought to be x. But that's crazy! What does it matter if at *some* time the chance of A is x? Surely we only have reason to set our credences equal to the chances if they're the *current* chances.

It's easy to slip into thinking about the chance-credence principle this way. But if we take this reasoning seriously, then we have to change PP$_1$: we have to time-index our credences, since the chance-credence link should only hold between our current credences and the current chances. This, of course, is how Vranas (2002) and Vranas (2004) understands the Principal Principle. And this is a mistake, for reasons we've already seen in section 2.2.2.

Here's another way to see how this reasoning is mistaken. Consider the above agent, but suppose that the one fact she knows is that the chance of A at t is x. Then PP$_1$ will say that this agent's credence in A ought to be x. And this is true regardless of what time the agent is located at, what time the agent believes she's located at, or even if the agent has no idea of what time it is. But then it's hard to see how knowing the time at which this chance fact obtains could possibly matter. If an agent has no idea what time it is, for example, it's hard to see how knowing that the chance of A is x *at* t gives her a reason to have a credence of x in A, while just knowing that the chance of A will be x at some unknown time does not.

Once we see how this reasoning is mistaken, it becomes clear that the time index that appears in PP$_1$ isn't really doing any work. So we can adopt a Principal Principle without time indexed chances—PP$_3$—and do just as well.

2.3.2 The Principal Principle and Admissibility

Perhaps the most puzzling feature of Lewis's (1986) paper is the fact that one of the versions of the Principal Principle he presents (PP_1) has an admissibility clause, while the other (PP_2) does not. If a satisfactory formulation of a chance-credence principle requires an admissibility clause, then presumably both should have one. If not, presumably neither should have one. This discrepancy is made stranger by Lewis's explicit ambivalence about spelling out what counts as admissible evidence: "I have no definition of admissibility to offer, but must be content to suggest sufficient (or almost sufficient) conditions..."[14] If PP_1 and PP_2 are equivalent, and PP_2 requires no admissibility clause, then providing a definition of admissibility should be trivial: call evidence "admissible" iff doing so makes PP_1 and PP_2 yield the same results.

The mainstream stance is that any satisfactory chance-credence principle requires an admissibility clause. On this view, Lewis's mistake was in his presentation of PP_2. He should have presented PP_2 as something like this:

$$\text{PP}_2^+ : hp(A|THE) = ch_{TH}(A), \text{ if } THE \text{ is admissible at } t.^{15} \tag{2.13}$$

A second stance is that neither PP_1 nor PP_2 needs an admissibility clause. Lewis's mistake was bringing admissibility into the mix in the first place. He should have presented PP_1 as this:

$$\text{PP}_1^- : \; hp(A|\langle ch_t(A) = x\rangle) = x \tag{2.14}$$

The stance that admissibility is unnecessary has been defended by Hall (1994), Hall (2004) and Meacham (2005).

[14] Lewis (1986*b*), p.92.

[15] Where t is the time H is a history up to.

A third stance is that admissibility isn't required for every satisfactory chance-credence principle, but it is required for a tidy presentation of chance-credence principles like PP_1, which are formulated using $\langle ch_t(A) = x\rangle$ terms instead of grounding arguments. This would explain the why Lewis equips PP_1 with an admissibility clause, but not PP_2. Unfortunately, this story won't work. As we'll see below, PP_2 entails PP_1^-. So if Lewis is correct in thinking that PP_2 doesn't need an admissibility clause, then PP_1 doesn't need one either.

In what follows, I'll argue that the second stance is correct: admissibility isn't needed. First, I'll argue that PP_2 doesn't require an admissibility clause. Then I'll show that PP_2 entails PP_1^-. Likewise, PP_2 entails an admissibility-free version of PP_3: PP_3^-. Since PP_2 doesn't require admissibility, and PP_2 entails admissibility-free versions of PP_1 and PP_3, those formulations don't require admissibility either.

Why Admissibility Isn't Needed

Let's look at whether PP_2 requires an admissibility clause. If we add an admissibility clause to PP_2, we get PP_2^+:

$$\mathrm{PP}_2^+ : hp(A|THE) = ch_{TH}(A), \text{ if } THE \text{ is admissible at } t. \tag{2.15}$$

PP_2^+ is strictly stronger than PP_2. We get PP_2 as a special case of PP_2^+ when E is a tautology.[16] So if we're worried about PP_2, we should be worried that PP_2 isn't strong enough without an admissibility clause: it doesn't give us all the relations between chance and credence that we intuitively think should hold.

With a little thought, though, we can see that PP_2, together with Bayesianism, *is* strong enough to capture all of the relations between chance and credence that we think should hold. First I'll sketch why this is the case, then I'll go through a few examples.

[16] Recall, we're following Lewis (1986*b*) in taking TH to always be admissible.

We can divide our uncertainty into two kinds, uncertainty about the outcomes of chance events and uncertainty about other things. We should only expect PP_2 to have a bearing on our uncertainty about the outcomes of chance events. But once we eliminate our uncertainty about other things, PP_2 completely fixes our credences in the outcomes of chance events. So no admissibility clause is needed to strengthen PP_2. The only way we could make PP_2 stronger is to have it fix our credences in things other than the outcomes of chance events. And we don't want the chance-credence principle to be that strong!

To get a feel for this, let's look at a couple of cases. First, let's look at an example of how PP_2 completely fixes our credences in the outcomes of chance events once we eliminate our uncertainty about other things. Assume, as we have been, that Lewis's metaphysical picture of chance (L1-L6) is correct. In particular, assume that chance distributions are functions of chance theories and complete histories up to a time. Consider a world where there are only two chance events, two fair coin flips, that take place at times t_1 and t_2, respectively. Consider a subject at this world who knows the laws T and the history up to t_0, H_0. Let T and H_0 entail everything about the world except how the coin tosses come up, so the subject knows everything about the world except the outcome of these chance events.

In this case the subject knows she's in one of 4 possible worlds. PP_2 and Bayesianism entail that her credence in each world should be $\frac{1}{4}$. If the subject learns the history up to t_1, H_1, then she'll be left with 2 worlds, and PP_2 and Bayesianism will entail that her credence in each of these remaining worlds should be $\frac{1}{2}$. Now consider the case we were worried about: what if she gets evidence such that her total evidence E consists of the laws and a *partial* history up to t_1? For example, what if she gets evidence that the coin won't land tails both times?

Once we're told what the new evidence is, it's simple to determine what her new credences should be. She should set her credence in the worlds incompatible with the evidence to 0, and normalize her credences in the rest. That is, like any good Bayesian, she should conditionalize. Since her credence in each of the 4 worlds was $\frac{1}{4}$, and her new evidence just eliminates the world where the coin lands heads twice, her new credence in each world should be $\frac{1}{3}$.

So if we eliminate our uncertainty about everything but the outcomes of chance events, PP_2 and Bayesianism completely fix our credences. The flip side of this is that if we make PP_2 stronger, it will constrain our credences in things other than the outcomes of chance events. And this would make it stronger than we want it to be.

Let's look at a case that illustrates the latter point. To do so we need to shelve L3, Lewis's claim that a chance distribution assigns chances to every proposition. (We can do this in good faith because, as we'll see shortly, this claim should be rejected.) Now consider a case like the case above: a world where there are only two chance events that take place at times t_1 and t_2, and a subject at this world who knows the laws T and the history up to t_0, H_0. In this case, though, T and H_0 *don't* entail everything about the world except how the coin tosses come up. T, H_0 and the outcomes of the coin tosses still leave certain features of the world unspecified. There might be, say, certain free willed spirits at this world, whose movements are neither chancy nor determined by anything in the past.

Now, say the subject learns that if the first coin toss comes up heads, the free willed spirits will all move to the left. What should her credences in the outcome of the first coin toss now be? This will depend on how much of her prior credence in heads is assigned to worlds where the spirits all move left. If all of it is assigned to such worlds, then her credence in heads/tails will remain the same. On the other hand, if her prior credence in such a possibility was 0, then her credence in heads/tails will become 0/1.

In this case, PP_2 fails to completely fix the subject's credences in the outcomes of chance events. But this is just as it should be. In order for PP_2 to fix the subject's credences in heads/tails in this case, it would have to tell us what credences to have in the movements of the spirits. And since the movements of these spirits have nothing to do with these chances, this isn't something that PP_2 should be telling us anything about.

Removing the admissibility clause from PP$_1$

So Lewis's formulation of PP_2 is correct: it doesn't need an admissibility clause. Given L1-L4, we can derive PP_1^- and PP_3^- from PP_2 (see Appendix 9.5). Since PP_2 is correct, it follows that PP_1^- and PP_3^- are correct as well. So none of the formulations of the Principal Principle require an admissibility clause.

It's worth noting that once we've disposed of admissibility, the three formulations of Principal Principle are no longer of the same strength. While we can derive PP_1^- and PP_3^- from PP_2 using only L1-L4, we cannot derive PP_2 from PP_1^- or PP_3^- without further assumptions. This is what we should expect. Of the three, PP_2 is the most fine-grained: it constrains priors conditional on the grounds of individual chance assignments. PP_1^- imposes less fine-grained constraints: it only constrains priors conditional on large unions of grounds which assign the same value to some proposition, and which are alike with respect to the time. And PP_3^- imposes even less fine-grained constraints than PP_1^-: it only constrains priors conditional on large unions of grounds that assign the same value to some proposition. So it should be no surprise that PP_2 places a stronger constraint on our priors than PP_1^- or PP_3^-.

The preceding discussion raises a natural question. If this is right, and chance-credence principles don't need admissibility clauses, why have so many people been mislead into thinking they do? Let's go through and examine two of the culprits: Miller's Principle and crystal ball cases.

Miller's Principle

Although Lewis's Principal Principle is the best known chance-credence principle, it was not the first. The first simple formalization of such a principle was offered by Miller (1966): "One's subjective probability for an event A, on supposition that the objective probability of A is x, ought to be x." I.e.,

$$\text{MP: } cr(A|\langle ch(A) = x\rangle) = x \tag{2.16}$$

But this constraint isn't entirely satisfactory. In particular, it doesn't seem like this constraint on our credences should always hold. Suppose we know that $\langle ch(A) = x\rangle$, but then learn A has in fact occurred. Our credence in $\langle ch(A) = x\rangle$ will remain 1 (we still know that our world grounds a chance distribution which assigns a chance of x to A), and our credence in A will be 1, so our credence in $cr(A|\langle ch(A) = x\rangle)$ should be 1, not x. To accommodate this sort of thing, it seems we need to add something further. And an admissibility clause seems to be just the thing we need: we should only take this principle to apply if we're not in possession of inadmissible evidence. This is a natural train of thought, and it has lead a number of authors, including Strevens (1995) and Vranas (2004), to conclude that an admissibility clause is mandatory.

These authors are certainly correct in thinking that Miller's Principle is not satisfactory. We have already seen reasons why principles of this form aren't tenable in section 2.2.2: conditional credence principles such as MP are generally inconsistent. What we want is a principle which constrains your credence in a proposition given that your total evidence is $\langle ch(A) = x\rangle$, not a principle that constraints your conditional credence regardless of what your evidence is. And if we follow Lewis and replace the conditional credences in Miller's Principle with conditional priors, or, alternatively, replace the conditional credences $cr(A|B)$ with $cr_B(A)$, then no inconsistencies arise.

These conditional credence formulations of chance-credence principles are at the root of much of the recent insistence that admissibility is required. (This doesn't, however, explain why Lewis (1986*b*) thought admissibility was needed.) For example, Vranas (2004) argues that the attempt in Hall (1994) to dispose of admissibility is misguided by raising the following objection: "Assume you know (for sure) that $ch(A)$ is 50%. Does it follow that your credence in A should be 50%?... Given [the chance-credence principle without an admissibility clause] it does, but in fact it doesn't, because you may also know some inadmissible proposition"[17], such as that A occurred.

If one understands the chance-credence principles to be claims about conditional credences, this objection makes sense. If your credences are bound by the constraint that $cr(A|\langle ch(A) = x\rangle) = x$, and your evidence includes $\langle ch(A) = x\rangle$, then your credence in A will have to be x, regardless of what your evidence is. But this is not the constraint Hall or Lewis are proposing. Hall and Lewis adopt principles that constrain one's conditional priors, not one's conditional credences. And these principles are not subject to the kind of worry Vranas raises, because they do not have the consequences he attributes to them. They only entail that a subject's credence in A should be $\frac{1}{2}$ if her *total* evidence is $\langle ch(A) = \frac{1}{2}\rangle$. If the subject is in possession of further evidence as well, such as that A has occurred, the principle no longer requires her credence in A to be $\frac{1}{2}$.

Crystal Balls

A number of people have considered the problem of how to apply chance-credence principles to crystal ball cases. These cases generally involve the following elements: a coin flip (at t_2, say), a crystal ball that suggests at some earlier time (t_0) how the coin will land, and a chance theory which assigns a non-trivial chance to the outcome of the coin flip (at t_1). The salient questions raised by such cases

[17] Vranas (2004), p.9

are these: If the subject knows only the information provided by the crystal ball and the grounds of the t_1 chance distribution, what should her credences be in the outcome of the coin toss? And how can this response be accommodated by one's favorite chance-credence principle?

The standard answer to the first question is that the subject's credences should no longer line up with the chances. But this answer appears to break with the advice of the chance-credence principle. This highlights the second question: how can this answer be reconciled with a chance-credence principle? The standard solution is to resort to admissibility: the evidence the crystal ball gives you is inadmissible, and so the chance-credence relation the principle would normally recommend is severed. Since this solution is not available to chance-credence principles without admissibility clauses, these cases have been taken to demonstrate the need for an admissibility clause.

This is a mistake: admissibility is not needed. So how should these crystal ball cases be treated? The correct analysis of crystal ball cases depends on the chance theory in question, and how we're thinking the crystal ball works. Hall (1994) describes why this is the case for one way of setting up these cases. Here's a catalog of how things go for several other set-ups as well.

1. *First variation:* Assume the crystal ball is infallible, and, following Lewis, let the grounds of chance distributions be chance theory T and complete history H pairs.

   This case is impossible. As Lewis showed, anything entailed by the arguments of a chance distribution will be assigned a chance of 0 or 1 by that distribution. (If $TH \Rightarrow A$, then $ch_{TH}(A) = hp(A|TH) = \frac{hp(ATH)}{hp(TH)} = \frac{hp(TH)}{hp(TH)} = 1$; if $TH \Rightarrow \neg A$, then $ch_{TH}(A) = hp(A|TH) = \frac{hp(ATH)}{hp(TH)} = 0$.) Since H includes the crystal ball result, and the crystal ball result entails the outcome of the coin toss, the distribution grounded by H, ch_{TH}, cannot assign a non-trivial chance to the outcome of the coin toss.

2. *Second variation:* Assume the crystal ball is infallible, and let the grounds of the chance distribution be a chance theory T and, say, some local features of the world L (i.e., the physical status of the coin flipping device, its immediate environment, and so on).

   In this case the crystal ball case is intelligible. How are your credences constrained if your total evidence is TLC (C being the information provided by the crystal ball)? It depends.

   (a) If the chance theory in question provides a chance distribution grounded by TLC, then your credence should accord with that chance. Since TLC entails the outcome of the coin toss, this chance is trivial (1 or 0).

   (b) Say the chance theory in question doesn't provide such a distribution. Then your credences will be what you'd get by lining them up with the chances grounded by TL, and then conditionalizing on C. When you conditionalize on C, you will rule out all of the possibilities in which the coin lands other than what the crystal ball predicted, since the crystal ball is infallible. So your credences in the outcomes of the coin toss will be 1 (for the result that the crystal ball indicates) and 0 (for the other result).

3. *Third variation:* Assume the crystal ball is fallible, and let the grounds of the chance distribution be T and H.

   What should your credences be if your total evidence is $THC = TH$? It depends on how we're thinking about the crystal ball. Presumably the crystal ball is of interest because we believe the crystal ball to be a "fallible but reliable guide to the future". We can understand "reliable" in two ways.

(a) We believe the crystal ball is "reliable" because it displays its result via a chance process which has a high chance of indicating the correct result, then this chance process will be part of the chance theory, T. In this case your credence will line up with the chance assigned by the chance distribution grounded by TH. (Here we need an explicit theory in hand to figure out what your credences will be.)

(b) The crystal ball's display is not the result of a chance process, but we believe the crystal ball is "reliable" because our priors in its being an accurate predictor are high. We can generate our credences by lining them up with the chances grounded by TH and then conditionalizing on C. (Here we need an explicit description of your priors to figure out what your credences will be.)

4. *Fourth variation:* Assume the crystal ball is fallible, and let the grounds of the chance distribution be T and L.

   Things here will proceed in the same way as in the third case.

2.3.3 Evaluating Lewis's Account of the Status of the Chance Credence Principle

Now let's turn to Lewis's claims about the status of the Principal Principle.

Does the Principal Principle Provide an Analysis of Chance?

Lewis (1986*b*) spends some time assessing the prospects for providing an analysis of chance using the Principal Principle. The idea is to simply define "$ch_{TH}(A)$" as the credence in A that you ought to have if your total evidence is TH.[18] This proposal runs into two immediate worries. First, the analysis identifies the chances grounded by TH as the credences you ought to have if your total

[18] Alternatively, instead of taking this to be a definition, we could take it to just be an analysis of what chances are, along the lines of water = H_2O.

evidence was TH. But this requires us to have an independent grasp of what T—a complete chance theory—is prior to knowing what chances are. Second, the analysis requires us to have an independent way of figuring out what our credences ought to be when our evidence is TH. In each case, Lewis tentatively suggests ways of salvaging the analysis.

In response to the first worry, Lewis suggests employing a Humean account of what a chance theory is. If we can characterize what chance theories are in Humean terms, then we'll have the independent grasp on them that we need in order to make the analysis non-circular.

I find this to be an odd response. If we can provide a Humean account of what chance theories are, then it seems we can use chance theories directly to provide an analysis of what chances are: chances are just the things that play the appropriate role in the chance theories. But why, then, are we trying to provide an analysis of chance using the Principal Principle? Once we have a Humean account of what chance theories are, then with respect to an analysis of chance, the Principal Principle looks superfluous.

In response to the second worry, Lewis suggests employing other principles or rationality, such as Indifference Principles, to try to cash out what our credence ought to be when our evidence is TH.

Again, this strikes me as an odd response. First, this option won't be available to Bayesians who take the only constraint on one's priors to be the constraint provided by the Principal Principle (a reasonably common position). Second, even if we were able to come up with consistent indifference principles which did constrain our credences in the relevant cases, there doesn't seem to be any reason other than blind faith to believe that the values they'll end yielding will line up with the "chances" we were talking about when we started, the chances that appear in our physical theories.

One might think that an analysis of chance in terms of other metaphysical primitives, like causal facts and counterfactuals, is possible, or one might think that a Humean analysis of chance is in reach. But in either case, it doesn't look like the Principal Principle will have much to do with it.

Does the Principal Principle Tell Us Everything We Know About Chance?

Lewis (1986*b*) claims that the Principal Principle captures everything we know about chance. Unlike other parts of Lewis's paper, this claim has received a great deal of attention, both critical (Arntzenius and Hall (2003)) and complimentary (Wallace (2006), and other proponents of the Many-Worlds interpretation of quantum mechanics). That said, I think this claim is often misunderstood. Upon reflection, I think Lewis isn't claiming anything as surprising or revolutionary as it first might seem. To see why, let's go through three ways of making sense of Lewis's claim, and assess the plausibility of each.

First, Lewis might mean that the only thing we know about chance is that they're things of which the following claim is true: the extent to which we believe a proposition should be proportional to the chance of that proposition, unless we're in possession of inadmissible evidence. This is the only thing we can make use of when we're evaluating whether or not something plays the chance-role well enough to deserve the name "chance".

This would be a surprising and interesting claim. But this claim is both clearly false and clearly not what Lewis intended. The claim that "the extent to which we believe a proposition should be proportional to the chance of that proposition" doesn't capture any of the six claims Lewis makes about chances (L1-L6). It doesn't entail that the objects of chance are *de dicto* propositions, or that the grounds of chances are chance theory and complete history pairs, or that determinism and chance are incompatible, and so on. Since Lewis clearly thought we do know these things about chance, this can't be what Lewis had in mind.

Second, Lewis might mean that we can formulate a relation between credence and chance that captures everything we know about chance. But this claim is trivially true. By adding conditions and constraints on the relata of the chance-credence relation we're describing, we can encode as much information about chance as we like. But this claim is no more interesting then the claim that all of the normative truths of the world can be captured by a single principle: the conjunction of all of the normative truths. And Lewis surely intended to say something more interesting than that.

Third, Lewis might have meant that the Principal Principle, as he formulates it, "packs in" enough constraints and conditions on the relata to capture everything we know about chance. If true, this would be interesting, and I think this is what Lewis intended to claim. That said, this claim is hard to evaluate, since it's not clear how much we should consider to be packed into the Principal Principle. If we allow him only what he states in the paragraph of exposition in which he introduces the principle (p.87), then even Lewis would take the claim is false, since it does not entail anything about the grounds of chance distributions, or that determinism and chance are incompatible, etc. If, on the other hand, we take the entire paper (and post-script) to be a prolonged exposition of the Principal Principle and its relata, then Lewis would probably take the claim to be true.

Whether *we* should take this claim to be true, however, is another question. Arntzenius and Hall (2003) have recently argued against Lewis's claim. They note, for example, that it seems the chance of a uranium decay of an atom in a certain state remains the same over time, and does not vary with the frequency of previous radioactive decays. But this is not entailed by Lewis's Principal

Principle, so the Principal Principle cannot encode everything we know about chance.

In any case, if this claim is true by Lewis's lights, then his formulation of the Principal Principle entails all of L1-L6. In the following sections I will argue that most of these claims are false. This gives us another reason to reject Lewis's claim: the Principal Principle cannot entail everything we know about chance if much of what it entails is false.

(Before moving on, it's interesting to note that this first way of understanding Lewis's claim has been adopted by recent Many Worlds proponents such as Wallace (2006), who use it to argue that the squared amplitudes in the Many Worlds theory are chances. First they approvingly cite Lewis's claim that the Principal Principle tells us everything we know about chance, and note that it follows from this that the only criterion we have for evaluating whether something is a chance is whether it satisfies the Principal Principle. Then they argue that our credences should be tied to the squared amplitudes in the same way as the Principal Principle ties our credences to the chances. Finally, they conclude that we have as much reason to take squared amplitudes to be chances as anything else. The fact that, say, the objects of square amplitudes aren't *de dicto* propositions isn't seen as a reason to think the square amplitudes aren't chances, because "the extent to which we believe a proposition should be proportional to the chance of that proposition" doesn't say anything about *de dicto* propositions. And since that's all we know about chance, we don't know that the objects of chances aren't *de dicto* propositions.

But, as we've seen, they are mistaken when they claim that their proposal is supported by Lewis. This first way of understanding Lewis's claim is not what he intended. And the other two ways of understanding Lewis's claim would not support their proposal.)

2.4 Revising the Chance-Credence Principle

What is the correct chance-credence principle?

I've argued that PP_2 is essentially the right chance-credence principle. But two wrinkles remain: PP_2 builds in Lewis's assumptions about the grounds of chance distributions and the propositions chances are assigned to. Since I'll argue that both of these assumptions are incorrect, it would be preferable to have a more general principle which doesn't incorporate these assumptions.

We can get such a principle as follows. Let G and A be any propositions. Then rational agents should satisfy the following constraint:

$$hp(A|G) = ch_G(A), \text{ if } ch_G(A) \text{ is defined.} \tag{2.17}$$

Using Bayesianism to formulate this directly in terms of credences, this becomes

$$\text{BP} : cr_G(A) = ch_G(A), \text{ if } ch_G(A) \text{ is defined,} \tag{2.18}$$

which I take to be the most general and elegant formulation of the chance-credence principle. I'll call this the *Basic Principle*.

Why have I chosen to present this principle in a form analogous to PP_2, instead of PP_1^- or PP_3^-? We've seen one reason already: once we remove the obscuring veil of admissibility, it becomes clear that PP_2-type constraints are more general than PP_1^- and PP_3^--type constraints. Another, more important, reason is that once we relax Lewis's constraints on what the grounds of chance distributions can be, the generalized analogs of PP_1^- and PP_3^-,

$$PP_1^G : cr_{\langle ch_t(A)=x \rangle}(A) = x, \text{ if } \langle ch_t(A) = x \rangle \neq \emptyset \tag{2.19}$$

$$PP_3^G : cr_{\langle ch(A)=x \rangle}(A) = x, \text{ if } \langle ch(A) = x \rangle \neq \emptyset, \tag{2.20}$$

won't generally hold.

If the grounds of chance distributions are such that there are never partial overlaps—i.e., G_i overlaps with G_j iff one is a subset of the other—then BP will entail PP_3^G. And if the grounds of chance distributions can also be time indexed in some natural way, then BP will entail PP_1^G as well.[19] Since both of these conditions hold if we take the grounds to be chance theory and completely history pairs TH, BP entails PP_1^G and PP_3^G if we adopt Lewis's account of the grounds of chance distributions. But if we allow grounds which do not satisfy these requirements, PP_1^G and PP_3^G won't always be true, and BP will not entail them. And we don't need to resort to strange and outlandish chance theories to find grounds that fail to satisfy these requirements. If we take statistical mechanics to be a chance theory, then statistical mechanics is one of them.

To get a feel for why PP_1^G and PP_3^G fail, let's go through an example. Consider a chance theory T which holds at only three worlds, w_1, w_2 and w_3. Let every subset of $\{w_1, w_2, w_3\}$ be the grounds of a chance distribution, and let each distribution assign an equal chance to each of the worlds that ground the distribution. So the distribution grounded by $w_1 \vee w_2 \vee w_3$ will assign a chance of $\frac{1}{3}$ to each world, the distribution grounded by $w_1 \vee w_2$ will assign a chance of $\frac{1}{2}$ to w_1 and w_2, and so on. Finally, let's stipulate that no other chance theory assigns a positive chance to any of these three worlds.

Now consider what a subject's credence in w_1 should be if her total evidence is $T = w_1 \vee w_2 \vee w_3$. BP delivers the correct answer:

$$cr_{w_1 \vee w_2 \vee w_3}(w_1) = ch_{w_1 \vee w_2 \vee w_3}(w_1) = \frac{1}{3}. \tag{2.21}$$

What about PP_3^G? Well, $\langle ch(w_1) = \frac{1}{3} \rangle = w_1 \vee w_2 \vee w_3$, so PP_3^G entails that

$$cr_{w_1 \vee w_2 \vee w_3}(w_1) = cr_{\langle ch(w_1) = \frac{1}{3} \rangle}(w_1) = \frac{1}{3}. \tag{2.22}$$

[19] These derivations are virtually identical to those given in appendix 9.5.

But it's also the case that $\langle ch(w_1) = \frac{1}{2} \rangle = (w_1 \vee w_2) \vee (w_1 \vee w_3) = w_1 \vee w_2 \vee w_3$, so PP_3^G entails that

$$cr_{w_1 \vee w_2 \vee w_3}(w_1) = cr_{\langle ch(w_1)=\frac{1}{2} \rangle}(w_1) = \frac{1}{2}. \tag{2.23}$$

So PP_3^G is inconsistent: it requires that your credence in w_1 be both $\frac{1}{2}$ and $\frac{1}{3}$.

What about PP_1^G? Here the problem is deeper. When we allow for grounds which can't be time indexed in a natural way, we can no longer make sense of $ch_t(A) = x$. So PP_1^G won't even apply.

Of course, if grounds can be time indexed, and are such that there are no partial overlaps, both PP_1^G and PP_3^G will hold, and we can employ them as we did before. But BP applies even when PP_1^G and PP_3^G do not.

BP is the correct chance-credence principle. What is its status? I take the status of BP to be no different from the status of other epistemic norms, like conditionalization. BP is just a normative claim about what our credences should be like. It does not provide an analysis of chance and it doesn't capture everything we know about chance.

2.5 Evaluating Lewis's Metaphysical Account

Now let's turn to Lewis's metaphysical account of chance. Lewis makes 6 claims, L1-L6, about the metaphysical structure of chance. Given our folk notion of chance, these claims look reasonable. But when we look at the chance theories entertained by physicists, many of these claims start to look implausible. Some of these claims are incompatible with particular kinds of chance theories: L4-L6 are incompatible with theories like classical and quantum statistical mechanics. And some of these claims, such as L2 and L3, raise worries with respect to pretty much every chance theory physicists have entertained. All of these problems are easier to evaluate once we have some concrete chance theories in hand. So I'll start by looking at the problems Lewis's account has accommodating particular kinds of chance theories, and then turn to the more general worries for his account.

2.5.1 Statistical Mechanics

Statistical mechanical theories pose a problem for Lewis's theory of chance. I'll draw out some of these problems by looking at a particular statistical mechanical theory, classical statistical mechanics. Since it will be useful to have a concrete theory to work with in the rest of this paper, I will sketch the theory in some detail.

Statistical Mechanics: A Sketch

At the foundation of classical statistical mechanics is classical mechanics. The world, according to classical mechanics, can be described like this. First, there are certain structural properties of the world. The structural properties include the spatiotemporal structure, the number of particles, and the intrinsic properties of these particles, such as their masses, charges, and so on. (For simplicity, I'm assuming that the only objects in the world are point particles.) Then there are the dynamic properties of the world. The dynamic properties we'll be concerned with are the positions of the particles and the rates of change of those positions. There are more dynamic properties than these: there are also the rates of change of the rates of changes of these positions, and so on, but these properties will end up being fixed by the properties already given.

Classical mechanics is deterministic, in the following sense.[20] Assume we know what the structural properties of the world are. If we're then given the dynamical

[20] Or rather, classical mechanics is *almost* deterministic. Earman (1986), Xia (1992), Norton (2003) and others have constructed classical mechanical cases in which determinacy fails. There are reasons, however, to think that this failures will have little impact on statistical mechanics. Although no proof of this exists, prevailing opinion is that these indeterministic cases form a set of Lebesgue measure zero. If so, we can ignore these cases in this context, since their exclusion will have no effect on the probabilities classical statistical mechanics assigns.

properties of the world at one time, we can predict the dynamical properties of the world at every other time. That is, if we're told what the positions and velocities of all of the particles are at one time, we can predict what they'll be at every other time.

Statistical mechanics makes use of a particular way of encoding the state of the world. Fix the structural properties of the world. Once we've done this, we can construct a *phase space* that corresponds to all of the possible dynamical properties—all of the possible particle positions and velocities—that are compatible with the structural properties we've fixed.[21] This phase space is usually represented as a 6N dimensional space, where N is the number of particles. For each given particle, three dimensions of the space are used to fix the three spatial coordinates of the particle, and three dimensions are used to fix the velocities of the particle in the three directions. The state of the world at a time can then be picked out by specifying a point in phase space. This point encodes the positions and velocities of every particle.

Since classical mechanics is deterministic, if we know the state of the system at one time, we'll know its state at every time. In other words, each point uniquely determines a path through phase space along which the system will travel, as the positions and velocities of the particles in the system change over time.

If we identify the point in phase space that corresponds to the world, then there isn't much more left to do. We can use the laws and the information we have to figure out anything about the world that we'd like to know. Of course, we don't usually know which point in phase space corresponds to our world. Instead, we have to make do with a rough, macroscopic description of the world. This macroscopic description won't be detailed enough to pick out a single point

[21] In versions of classical statistical mechanics like that proposed by Albert (2001), one of the statistical mechanical laws is a constraint on the initial entropy of the universe. On such theories the space of classical statistical mechanical worlds (and the phase spaces that partition it) will only contain worlds whose initial macroconditions are of a suitably low entropy.

in phase space, of course, since there will be many particle positions and velocities compatible with such a description. But it will pick out a region of phase space in which our world is located.

This is the domain of statistical mechanics. Call a single point in phase space a *microstate*, and a region of phase space (usually corresponding to a given macroscopic description) a *macrostate*. Statistical mechanics provides probabilities for the world being in one macrostate, given that the world is some other macrostate. Let m be the Liouville measure, the Lebesgue measure over the canonical representation of the phase space, and let A and B be two macrostates. Then the classical statistical mechanical probability of A given B is $\frac{m(A \cap B)}{m(B)}$.

Here is an example. Suppose an ice cube is placed in a cup of hot water. Then statistical mechanics assigns a probability to it melting in the following way.[22] Take the Liouville measure of the region of phase space occupied by microstates where a particular cup of hot water contains an ice cube, and the Liouville measure of the set of microstates in this region in which the ice cube melts, and then divide the later by the former. Since the overwhelming majority of these microstates, according to the Liouville measure, are ones in which the ice cube melts, statistical mechanics yields the verdict that it's overwhelmingly likely that the ice cube will melt.

Note that statistical mechanical probabilities aren't defined for all object propositions A and background state propositions B. Given the above formula, two conditions must be satisfied for the chance of A relative to B to be defined: both $m(A \cap B)$ and $m(B)$ must be defined, and the ratio of $m(A \cap B)$ to $m(B)$ must be defined.

Despite the superficial similarity, the statistical mechanical probability of A relative to B is not a conditional probability. If it were, we could define the

[22] Assuming, of course, that we have a way of precisely spelling out what particle configurations satisfy these macroscopic descriptions.

probability of A '*simpliciter*' as $m(A)$, and retrieve the formula for the probability of A relative to B using the definition of conditional probability. The reason we can't do this is that the Liouville measure m is not a probability measure; unlike probability measures, there is no upper bound on the value a Liouville measure can take. We only obtain a probability distribution after we take the ratio of $m(A \cap B)$ and $m(B)$; since $m(A \cap B) \leq m(B)$, the ratio of the two terms will always fall in the range of acceptable values, [0,1].

Lewis's Account and Statistical Mechanical Chances

Statistical mechanical probabilities cannot be chances on the Lewis's account. As we'll see below, there are general worries about whether physical chance theories, including classical statistical mechanics, could satisfy L2 and L3. But statistical mechanics in particular poses further difficulties for Lewis's account of chance.

First, classical statistical mechanical chances are compatible with classical mechanics, a deterministic theory. But according to L6, determinism and chance are incompatible. So statistical mechanics is incompatible with L6.

Second, L4 and L5 are incompatible with statistical mechanical chances. L4 requires that the grounds of chance distributions are conjunctions of chance theories and histories up to a time. L5 requires that anything entailed by the history that grounds a distribution must be assigned a chance of 1 by that distribution. Since every history entails at least the initial conditions, L4 and L5 together entail that chance distributions must assign trivial chances (1 or 0) to initial conditions. But there are statistical mechanical probability distributions which assign non-trivial probabilities to initial conditions. (Indeed, *every* non-trivial statistical mechanical probability distribution assigns non-trivial probabilities to initial conditions.) So statistical mechanical probabilities cannot be chances.

The contradiction just described employed both L4 and L5. If we want to accommodate statistical mechanics, can we reject just L5 and keep L4? Not if we accept something like PP_2, since L1-L4 and PP_2 entail L5 directly (see Appendix 9.2). How about rejecting L4 and keeping L5? This is a logical possibility, but a tricky one to carry out in a satisfactory manner. Once we reject L4, it's hard to make sense of the claim that the past isn't chancy. The claim that the past is no longer chancy presupposes that chance distributions can be associated with a time, relative to which some events are in the past. On Lewis's account chance distributions can be associated with a time because they're functions of chance laws and histories, and chance laws and histories can be picked out by a world and a time. But once we reject L4 we're no longer guaranteed a way to associate a time with chance distributions, and thus no longer guaranteed a way to make sense of the claim that the past is no longer chancy.

Why Statistical Mechanical Probabilities are Chances

How should we understand statistical mechanical probabilities? A satisfactory account must preserve their explanatory power and normative force. For example, classical mechanics has solutions where ice cubes grow larger when placed in hot water, as well as solutions where ice cubes melt when placed in hot water. Why is it that we only see ice cubes melt when placed in hot water? Statistical mechanics provides the standard explanation. When we look at systems of cups of hot water with ice cubes in them, we find that according to the Liouville measure the vast majority of them quickly develop into cups of lukewarm water, and only a few develop into cups of even hotter water with larger ice cubes. The explanation for why we always see ice cubes melt, then, is that it's *overwhelmingly likely* that they'll melt instead of grow, given the statistical mechanical probabilities. In addition to explanatory power, we take statistical mechanical probabilities to have normative force: it seems irrational to believe that ice cubes are likely to grow when placed in hot water.

These desiderata are met If we take statistical mechanical probabilities to be chances. Statistical mechanical probabilities have the explanatory power they do because they're chances; they represent lawful, empirical and contingent features of the world. Likewise, statistical mechanical probabilities have normative force because they're chances, and chances normatively constrain our credences via something like the Principal Principle.

But as we've just seen, Lewis's account is committed to denying that statistical mechanical probabilities are chances. The alternative is to take them to be subjective values of some kind. There's a long tradition of taking statistical mechanical probabilities to represent the degrees of belief a rational agent should have in a particular state of ignorance. It proceeds along the following lines.

Start with the intuition that some version of the Indifference Principle—the principle that you should have equal credences in possibilities you're epistemically 'indifferent' between—should be a constraint on the beliefs of rational beings. There are generally too many possibilities in statistical mechanical cases—an uncountably infinite number—to apply the standard Indifference Principle to. But given the intuition behind indifference, it seems we can adopt a modified version of the Indifference Principle: when faced with a continuum number of possibilities that you're epistemically indifferent between, your degrees of belief in these possibilities should match the values assigned to them by an appropriately uniform measure. The properties of the Lebesgue measure make it a natural candidate for this measure. Granting this, it seems the statistical mechanical probabilities fall out of principles of rationality: if you only know B about the world, then your credence that the world is in some set of states A should be equal to the proportion (according to the Lebesgue measure) of B states that are A states. Thus it seems we recover the normative force of statistical mechanical probabilities without having to posit chances.

However, as Albert (2001), Loewer (2001), and others have argued, this account of statistical mechanical probabilities is untenable. First, the account suffers from a technical problem. The representation of the state space determines the Lebesgue measure of a set of states, and there are an infinite number of ways to represent the state space. So there are an infinite number of ways to 'uniformly' assign credences to the space of possibilities. Classical statistical mechanics uses the Lebesgue measure over the canonical representation of the state space, the Liouville measure, but no compelling argument has been given for why *this* is the right way to represent the space of possibilities when we're trying to quantify our ignorance. So it doesn't seem that we can recover statistical mechanical probabilities from intuitions regarding indifference after all.

Second, the kinds of values this account provides can't play the explanatory role we take statistical mechanical probabilities to play. On this account statistical mechanical probabilities don't come from the laws. Rather, they're *a priori* necessary facts about what it's rational to believe when in a certain state of ignorance. But if these facts are *a priori* and *necessary*, they're incapable of explaining *a posteriori* and *contingent* facts about our world, like why ice cubes usually melt when placed in hot water. Furthermore, as a purely normative principle, the Indifference Principle isn't the kind of thing that could explain the success of statistical mechanics. Grant that *a priori* it's rational to believe that ice cubes will usually melt when placed in hot water: that does nothing to explain why in fact ice cubes *do* usually melt when placed in hot water.

So the indifference account of statistical mechanical probabilities is untenable. Since the only viable account of statistical mechanical probabilities on offer is that they are chances, and Lewis's account is incompatible with statistical mechanical chances, Lewis's account needs to be revised.

A Hitch: Statistical Mechanics and the Chance-Credence Relation

Before we move on to the other worries for Lewis's account, let me mention a potential difficulty that statistical mechanics raises for chance-credence principles formulated in terms of hypothetical priors, like PP_2 or the hypothetical prior formulation of BP, (2.17). The problem is that on PP_2 and (2.17) our priors for theories like classical statistical mechanics are constrained only by trivial chances. And so the values of the non-trivial statistical mechanical chances are epistemically irrelevant, since they have no effect on our priors.

A rigorous derivation of this result is given in Appendix 9.6, but the following is a rough sketch of how the problem arises. If the background state B of a statistical mechanical chance is of infinite measure, then that chance will be trivial or undefined.[23] So the background state B of a non-trivial chance must be of finite measure. Now, any prior you have in a classical mechanical state space is required by the chances to be spread uniformly over that space in accordance with the Liouville measure. Since the state spaces of classical statistical mechanics are of infinite measure, any finite measure region of such a space will be assigned a 0 prior.[24] So the background state B of a non-trivial chance will be assigned a 0 prior. But the Basic Principle only applies if one's prior in the arguments TB of the chance distribution are non-zero, since otherwise $hp(A|TB)$ is undefined. Since one's prior in the background state B of any non-trivial chance will be 0, it follows that the Basic Principle never applies to non-trivial statistical mechanical chances.

We saw the source of the problem in our sketch of statistical mechanics. The problem arises because chance-credence principles like PP_2 and (2.17) attempt to

[23] I'm assuming the extended real number line and the standard extension of the arithmetical operators over it; in particular, that $\frac{x}{\infty} = 0$ if x is finite, and $\frac{\infty}{\infty}$ and $\frac{x}{0}$ are undefined.

[24] There is one state space of finite measure, the trivial state space of a system with no particles. But since the chances associated with this space are trivial, we can safely ignore it.

equate statistical mechanical chances with conditional priors. But as we saw in section two, we can't equate the statistical mechanical chance of A relative to B with a conditional probability. To do so would require us to make sense of the probability of A *simpliciter*, where the probability of A *simpliciter* is set equal to the Liouville measure of A. But the Liouville measure is not a probability measure, since there is no upper bound to the values it can assign. So these values generally won't make sense as probabilities. The clauses in PP_2 and (2.17) that require $hp(A|TB)$ and $ch_{TB}(A)$ to be defined prevent contradictions by severing the chance-credence connection in problematic cases. But after severing the problematic chance-credence connections we find that most statistical mechanical chances don't have an effect on our priors, and those that do are trivial.

One way to respond to this problem is to adopt a chance-credence principle like the credence formulation of BP, that equates chances with credence-given-total-evidence. Since this principle doesn't attempt to equate chances with conditional probabilities, it avoids the problems that (2.17) runs into.

A second way to respond to this problem is to follow Hajek (2003) and reject the standard definition of conditional probabilities. Hajek proposes that we take conditional probabilities to be primitive, and understand the formula $p(A|B) = \frac{p(A \wedge B)}{p(B)}$ to be a constraint on the values of conditional probabilities when $p(B) > 0$. Adopting Hajek's proposal avoids the problem because $hp(B) = 0$ no longer entails that $hp(A|TB)$ is undefined, and thus (2.17) can still apply when $ch_{TB}(A)$ is non-trivial. If we adopt this response, we can still have something like (2.17) as our chance-credence principle. (Extensions of standard probability spaces more general than, such as lexicographic probability spaces and non-standard probability spaces, can also be employed.[25])

[25] See Halpern (2001) for a discussion of these accounts, and for proofs that they're strictly more general than primitive conditional probability spaces (Popper spaces).

So what should we conclude? If we adopt one of the hypothetical prior accounts of *de se* beliefs I discuss in chapter 5 then we'll have to adopt the second option. But other than that, nothing much hangs on which response we endorse. So I'll remain neutral between these two responses.

2.5.2 When Chance Distributions Are Defined

L3 states that chance distributions assign values to every proposition (or at least every proposition to which an idealized credence function assigns values). But this isn't generally true for the chance theories entertained in physics.[26]

As an example, consider classical statistical mechanics. Given the structural constraints and a background state, statistical mechanics assigns chances to particles having certain positions and velocities. And that's it. If there are ghostly spirits that don't supervene on the positions and velocities of particles, statistical mechanics won't assign chances to these ghosts being one way or another. If there were non-supervenient consciousness facts, statistical mechanics won't assign chances to these facts being one way or another. There's a limit to how detailed the possibilities are to which statistical mechanics assigns chances: they have to be possibilities you can cash out in terms of particle positions and velocities.

More generally, the kinds of chance theories entertained by physicists only assign chances to propositions that pertain to the features of the world relevant to the chance theory in question. So L3 is false: chance distributions don't assign values to every proposition.

[26] Hoefer (2005) has also argued that Lewis is mistaken on this point.

2.5.3 The Objects of Chance

Another potential problem for Lewis's account comes from L2, the claim that the objects of chance are *de dicto* propositions. The problem arises for chance theories whose models have certain physical symmetries.

Consider an example of this problem in classical statistical mechanics.[27] Take a classical statistical mechanical state space S. Consider two disjoint regions in S of finite and equal Liouville measure that are related by a symmetry transformation. That is, the points in the first region map to the points in the second by a rotation about a given axis, a spatial translation, or some other symmetry of the relevant systems. Let A_1 and A_2 be the first and second regions, and let B be the union of these regions. What is the statistical mechanical chance of A_1 relative to B? Since the Liouville measure of A_1 is half that of B, the chance of A_1 relative to B should be $\frac{1}{2}$. Likewise, the chance of B relative to B should be 1.

Now, the objects of statistical mechanical chances are regions of state space. According to L2, the objects of chances are *de dicto* propositions, i.e., sets of possible worlds. So it needs to be the case that we can take regions of state space to correspond to sets of possible worlds. In situations with symmetries like the one sketched above, it's hard to see what set of worlds to associate with a region of state space like A_1. The worlds in A_1 are qualitatively identical to the worlds in A_2, and qualitatively identical worlds are generally thought to be numerically identical. So if we say A_1 contains a world if any of its state space points correspond to that world, then it will contain the same worlds as A_2 and B. But if A_1 and B are the same proposition, then the chance of A_1 relative to B should be the same as the chance of B relative to B, which it is not.

[27] Cases of this form have been discussed in the context of the hole argument by Wilson (1993) and Arntzenius (2003*b*).

Alternatively, if we say A_1 contains a world if it contains all of the state space points that correspond to that world, then A_1 will contain no worlds. But if A_1 is the empty set, then it follows from the probability axioms that $ch_{TB}(A_1) = 0$, which it does not.[28]

The problem stems from the tension between three individually plausible assumptions. The first assumption is that our chance theories successfully assign the chances they seem to assign. The second assumption is that there are no non-qualitative differences between possible worlds. This assumption addresses the intuitive difficulty of making sense of qualitatively identical but distinct possible worlds. The third assumption is that the objects of chances are *de dicto* propositions. This captures the intuition that chances are about the way the world could be. In these terms, the problem is that our chance theories seem to assign chances which are hard to make sense of if we take the objects of chance to be sets of possible worlds and take qualitatively identical worlds to be identical.

A natural response to this problem is to reject one of these three assumptions. One option is to reject the first assumption, and reject as unintelligible any apparent chance assignments whose objects or arguments don't neatly correspond to sets of possible worlds. In the context of classical statistical mechanics, this constraint will be that the object and background state of a chance assignment must contain either all of the state space points corresponding to a world or none of them. In the above example, this gets around the problem of making sense of the chance of A_1 relative to B by denying that such chances are intelligible.

Another option is to reject the second assumption, and use *haecceities* to individuate between qualitatively identical worlds. With *haecceities* we can distinguish between worlds related by symmetry transformations, and make sense of chances with these worlds as objects. In the above example, this makes analyzing

[28] That $ch_G(\cdot)$ is a probablity function over possible worlds follows from the criteria laid out in section four and the assumption that the objects of chances are sets of possible worlds.

the chance of A_1 relative to B straightforward, since A_1 and B represent distinct and well-defined sets of possible worlds.

A third option is to reject the third assumption and take the objects of chances to be something other than (*de dicto*) propositions. On this approach A_1 would not correspond to a set of possible worlds; instead, the chance of A_1 relative to B would be made intelligible by resorting to an alternative account of the relevant space of possibilities. This option is more open ended then the first two. In addition to providing a different account of the objects of chance, this response requires a different chance-credence principle. Chance-credence principles like the Basic Principle and the Principal Principle equate values associated with the same objects; i.e., they equate the chance of a *de dicto* proposition with a credence in it. Changing the objects of chance from *de dicto* propositions to Xs requires a modification of the chance-credence principle to account for this. Either the chance-credence principle must be modified to account for how chances in Xs link up with credences in *de dicto* propositions, or the chance-credence principle must be modified so that chances in Xs are linked up with credences in Xs, and an account of credence in these Xs must be provided.

So what should we conclude? If we adopt the first two options then we can hold on to L2. If we adopt the third option, then we must reject L2. I'm inclined to keep L2, so I'm inclined to adopt one of the first two options. I think that L2 captures something central to the concept of chance: that the objects of chance, what chances are about, is ways the world could be. This is why, I propose, we find it so conceptually difficult to make sense of the squared amplitudes of the Many Worlds interpretation as chances. On the Many Worlds theory the intensities attach to different branches—different parts of the same world—not to different possible worlds. And chances apply to different ways the world might be, not to different branches in a world that will be.

2.6 Revising the Metaphysics of Chance: M1-M4

In the previous section we saw several problems with Lewis's metaphysical account of chance. Now I'll propose an alternative to Lewis's account which avoids these difficulties. Before we examine this alternative, however, let me briefly sketch the metaphysical big picture I'm presupposing.

In physics we find a bunch of theories that deal with objective probabilities: "chance theories". I take these theories, if they're true, to be part of the natural laws. I take the objective probabilities that they describe to form a natural kind: chances. So we can characterize chance theories as the parts of the laws which involve this natural kind, the chances.[29] In what follows, one may see my claims about chance as claims about metaphysical possibility: in all possible worlds, chances and chance theories have the structure that I will propose, and can be described in the manner I will suggest.

2.6.1 An Account of the Structure of Chance: M1-M3

My first three metaphysical claims about chance are the following (I'll make a fourth claim, regarding the structure of chance theories, in the next section):

M1. Every possible chance assignment can be encoded by a single function, *ch*, which takes a *grounding argument* G and spits out a probability function $ch_G(\cdot)$.

M2. Chances are assigned to ways the world could be; i.e., *de dicto* propositions. So the chance distributions $ch_G(\cdot)$ are probability functions over *de dicto* propositions.

[29] If one's picture of laws makes this statement ambiguous, one can simply throw out the notion of chance theories, and with a few tweaks, reframe everything I've said in terms of the laws that hold at worlds where chances are instantiated.

M3. The grounding argument (or *grounds*) of the chance distribution is a conjunction of a complete chance theory T and a background state, B. (What kinds of background states yield well-defined chance distributions depends on the chance theory in question.) Note that:

(i) T and B entail the chance distribution they ground.

(ii) Chance distributions supervene on chance theories T and background states B.

The first two claims are identical to L1 and L2. The third claim is similar to L4, but allows for a wider range of grounding arguments. Lewis's metaphysical account of chance was tailored to fit *dynamical chances.* This is because Lewis thought that all chances were dynamical chances. And many of his claims, such as L4-L6, have features that reflect this assumption. My counterexamples to L4-L6 employ non-dynamical chances, the chances of classical statistical mechanics. This raises some natural questions. I've rejected L4-L6 because they don't hold for non-dynamical chances. But do I think they still hold for dynamical chances? And to the extent to which I do, what are Lewis and I disagreeing about? Let me address each of these questions in turn.

L4-L6 are certainly more plausible when restricted to dynamical chances, but all of them are contestable.

I think L4 is questionable. It doesn't seem like the grounds of dynamical chance distributions need to be chance theory and history conjunctions. It seems plausible, for example, to take Markovian chance distributions to be grounded by the conjunction of a chance theory and a thin time slab consisting of the most recent states of the system. That said, it does seem plausible that there will always be a natural way to time index the grounds of dynamical chances distributions, as Lewis claimed.

I think L5 is questionable as well. First, if we no longer endorse L4, then Lewis's principled reason for endorsing L5 disappears. If the grounds of dynamical chance distributions are thin time slabs of prior temporal states (say) instead of complete histories, PP_2 will no longer entail that the past is no longer chancy. Second, L5 is unattractive if we want to allow for the possibility of time symmetric dynamical chance theories. Interestingly, Lewis was inclined to allow for this possibility, and viewed the temporal asymmetry built into his account as a deficit: "Any serious physicist, if he remains at least open-minded both about the shape of the cosmos and about the existence of chance processes, ought to do better. But I shall not..."[30])

What about L6? I think L6 is plausible for dynamical chances, though in this case much hangs on the notion of determinism employed, and the details of how one distinguishes between dynamical and non-dynamical chances.

On to the second question. Suppose I were to grant that L4-L6 apply to dynamical chances. Would Lewis and I still be disagreeing about anything important? We seem to diverge on whether statistical mechanical probabilities are chances, of course, but why does it matter if statistical mechanical probabilities are chances are not? Why do we need to take statistical mechanical probabilities to be chances?

Here is what I'm committed to with respect to statistical mechanical probabilities. (i) We need something like the BP to apply to statistical mechanical probabilities if we are to get the normative tie we want between these probabilities and our credences. (ii) We need statistical mechanical probabilities to play an explanatory role similar to that of dynamical chances. (iii) Statistical mechanical probabilities and dynamical chances are similar in a number of ways. M1-M3 hold of both of them, BP holds for both of them, and (as I'll propose below for

[30] Lewis (1986*b*), p. 94.

M4) the laws invoking them have the same general form.

Given this, it seems to me that statistical mechanical probabilities and dynamical chances are more alike than different. And it seems natural to classify them as different kinds of the same thing: chances. That said, as long as one agrees with (i)-(iii), I have no objection with calling only dynamical chances "chances". We can call statistical mechanical probabilities something else—schmances, say—and just understand the claims I'm making about chances (BP and M1-M4) to be claims about both chances and schmances.

Lewis, of course, would not be happy to concede this. He would reject all of (i)-(iii). So this is the substance of our disagreement.

2.6.2 An Account of the Structure of Chance Theories: M4

The three claims I made above largely pertain to the structure of the chance function *ch*. Now I will present a claim about the structure of chance theories.

The Structure of Chance Theories

M4. Every chance theory T has the following structure:

(i) The worlds where the theory holds can be partitioned into *coarse sets.*

(ii) Each coarse set can be partitioned into *fine sets.*

(iii) Each coarse set C is associated with a countably additive measure m_{TC}, which is defined on an algebra that includes all of the fine sets of C but no proper subsets of these sets except the empty set.

(iv) The chance of A given chance theory T and background state B is:

$$ch_{TB}(A) = \frac{m_{TC \supset B}(AB)}{m_{TC \supset B}(A)}, \tag{2.24}$$

where $m_{TC\supset B}$ is the measure associated with the coarse set containing B.

What do these bits of structure intuitively represent? The coarse sets correspond to the least detailed background state propositions relative to which the theory assigns well-defined chances. The fine sets correspond to the most detailed object propositions to which the theory assigns well-defined chances. The measures encode the chances of the theory, although they themselves need not be probability measures. And given this structure, (2.24) determines everything about the chances of T. It determines which chance distributions are well-defined for T, what grounds these distributions, and what values these distributions assign to which propositions.

In the case of classical statistical mechanics, the coarse sets are the phase spaces; i.e., sets of classical statistical mechanical worlds which share the relevant structural properties. The fine sets are the points of the state spaces; i.e., individual possible worlds. The measures are the Liouville measures over the phase spaces. Given this, (2.24) lines up with the chances that classical statistical mechanics assigns. In particular, note that $ch_{TB}(A)$ is well-defined iff (2.24) is defined, and so is defined iff (a) T is a complete chance theory and B is a subset of a coarse set C of T, (b) the ratio of $m_{TC\supset B}(A \cap B)$ to $m_{TC\supset B}(B)$ is defined, and (c) $A \cap B$ and B are elements of S, the algebra over which $m_{TC\supset B}$ is defined. As we saw in section 2.5.1, this lines up with the conditions under which classical statistical mechanical chances are defined.

Is M4 the constraint on chance theories we want to adopt? I thought so when I first proposed it, in Meacham (2005). Since then, however, I've come to think that M4 should be slightly modified. My worry with M4 stems from the possibility of chance theories with multiple layers of coarse sets and fine sets. When I wrote Meacham (2005), I thought that multiple layers would always be "glueable"—fine sets of one layer would be coarse sets of another—so one could always formulate chance theories in terms of a single coarse set and fine set partitioning. And, as far as I'm aware, this is true for all of the chance theories entertained in physics. (We'll look at a glueable multi-layer theory in the following section.) But there are reasonable-looking chance theories for which this is not the case.

For example, consider a theory which (a) assigns chances to there being a certain numbers of particles given the spatiotemporal extension of the world, and (b) assigns chances to the particles having particular positions and velocities given the spatiotemporal extension of the world, the number of particles, and the masses of these particles. The coarse set of the (a)-layer consists of sets of worlds with the same spatiotemporal extension, and the fine sets of each coarse set are sets of worlds which also have the same number of particles. The coarse sets of the (b)-layer consists of sets of worlds which are alike with respect to spatiotemporal extension, number of particles and particle masses, and the fine sets of each coarse set are sets of worlds which also have the same particle positions and velocities. In this case the (a)-layer and the (b)-layer have a gap between them—the fine sets of the (a)-layer are strictly larger than the coarse sets of the (b)-layer—so we can't glue them together and replace them with a single coarse set and fine set partition.

If we want to allow for chance theories like this, we need to tweak M4 to allow for several layers. For those interested in these details, they can be found in Appendix 9.7. That said, assessing the implications of M4 is much simpler if we restrict our attention to single layer chance theories, and none of the issues we're concerned with depend on this detail. So in what follows I'll assume that all of the chance theories in question are single layer chance theories.

A Retrospective: Reconsidering the Chance-Credence Principle in Light of M1-M4

With this alternative account of chance in hand, let's return to the question of how our priors should be constrained by the chances according to the Basic Principle. There may be a number of objective constraints on one's priors, but in this context we're only interested in those imposed by the chances. So, for simplicity, let us assume a version of subjective Bayesianism on which the Basic Principle is the only objective constraint on our priors.

Given the structure outlined above, we can divide up the space of possible worlds into smaller and smaller regions by applying finer and finer partitions. We can partition the space of possible worlds into sets of worlds where a given chance theory obtains, partition the worlds where a given chance theory holds into coarse sets, partition these coarse sets into fine sets, and (if needed) partition these fine sets into individual possible worlds. (Note that I'm taking non-chancy worlds to have a "chance theory" too—the trivial chance theory T with one coarse set and one fine set, and (thus) only trivial chance assignments: the chance of A is 1 if $A \supset T$, 0 if $A \cap T = \emptyset$, and undefined otherwise.)

Now, we're interested in how the chances constrain our priors. Our priors in all possibilities must sum up to 1. So the question of interest is this: what bearing do the chances have on how our priors get divided up? Let's see how to divide our priors in steps, following the partition structure just sketched above.

First we divide our priors up among the different chance theories that can hold. Since we need to assume that a particular chance theory holds before we can get any chances, how we should divide our priors among chance theories is beyond the scope of chance. So we divide our prior among chance theories subjectively.

Next we divide our prior in a given chance theory among its coarse sets. Since we need to fix on a coarse set before a theory can assign chances, how we should divide our prior in a chance theory among its coarse sets is also beyond the scope of chance. So we divide our prior among coarse sets subjectively.

Next, we divide our prior in a coarse set among its fine sets. This is where the chances come in. The chances restrict how our priors in coarse sets can be divided among our the fine sets. So we divide our priors among the fine sets in accordance with the chances.

Finally, if there are further features of the world that the chances don't pertain to, like (say) facts about ghosts, then we divide our prior in a fine set among its individual possible worlds. Since the fine sets are the smallest units to which chances are assigned, once we've fixed on a fine set the chances have nothing more to say. So we divide our prior among individual possible worlds subjectively.

And that's it. Now we can see exactly how our priors are restricted by the chances. Given M4, the chances rigidly determine how our prior in the coarse sets of a theory is divided among the fine sets of the theory. That's the entirety of it.

(More formally, we can express our hypothetical prior in an arbitrary proposition A in terms of the partitioning structure described above. So let T_i be the chance theories, C_j the coarse sets, F_k the fine sets and W_l the individual possible worlds. Then:

$$hp(A) = \sum_{i,j,k,l} hp(T_i)hp(C_j|T_i)hp(F_k|C_j)hp(W_l|F_k)hp(A|W_l) \tag{2.25}$$

All of these terms are determined subjectively except for $hp(F_k|C_j)$,[31] which is fixed by the chances.[32] (See Appendix 9.8 for details.))

[31] A caveat is required here, given certain kinds of infinity issues that arise (see section 2.5.1). Strictly speaking, these terms are sometime undefined. If this is the case, and the other constraints the chances impose don't fix the value of the term, its value will be determined subjectively.

[32] I'm implicitly assuming that the indices i, j, k, l range over countably infinite members at most.

Let's end by looking at how this applies to some of the chance theories of physics. How do the chances of classical statistical mechanics constrain our priors? To determine our priors in the classical statistical mechanical worlds, we subjectively determine our prior in classical statistical mechanics, divide this subjectively among the phase spaces, and divide our prior in each phase space among its points in accordance with the statistical mechanical chances. If the points of state space are individual possible worlds, we don't need to divide our priors any further; if not, we divide it among the worlds subjectively. So the chances of classical statistical mechanics constrain how our priors in phase spaces are assigned to the points of phase space.

Now let's look at how the proposal works for a different chance theory, statistical Bohmian mechanics. Statistical Bohmian mechanics is the complete chance theory encompassing Bohmian mechanics and quantum statistical mechanics. Unlike classical statistical mechanics, the chances of statistical Bohmian mechanics are generally segregated into the chances of Bohmian mechanics and the chances of quantum statistical mechanics. To apply the third proposal to statistical Bohmian mechanics we need to glue the chances of Bohmian mechanics and quantum statistical mechanics together, and fit them into the framework given above.

I will first give a brief description of quantum statistical mechanics and Bohmian mechanics. To avoid a lengthy discussion of these theories, I won't present them in as much detail as I presented classical statistical mechanics. Instead, I will simply give a gloss of their relevant features, and then sketch how each fits into the above framework.

As with classical statistical mechanics, quantum statistical mechanics starts with spaces of possibilities that share certain structural properties, such as spatiotemporal dimensions of the system, the number of particles, etc. The elements

Strictly speaking, this assumption should be discarded and these sums should be replaced by integrals over the appropriate probability densities.

of these spaces are picked out by certain dynamic properties, in this case the property of having the same wave function at a given time. Quantum statistical mechanics assigns a canonical measure over these possibilities, and this measure yields the chances.[33]

Bohmian mechanics is an interpretation of quantum mechanics that adds hidden variables to the formalism, in this case the positions of the particles. In Bohmian mechanics a complete description of a system at a time is given by the structural properties considered above, as well as the wave function and particle positions of the system. Both the wave function and the particles evolve deterministically, so a complete description of the system at a time fixes the history of the system. Bohmian chances come in when we consider possibilities that have the same wave function and relevant structural properties but differ in particle positions. Bohmian mechanics assigns a special measure over this space, and this measure yields the chances.[34]

The framework given above straightforwardly applies to each of these theories. In quantum statistical mechanics the coarse sets are sets of possibilities that share the relevant structural properties, and its fine sets are the sets of possibilities with the same wave function. In Bohmian mechanics the coarse sets are sets of possibilities that share the relevant structural properties and have the same wave function, and its fine sets are the sets of possibilities with the same particle positions. Since the relevant structural properties, wave function, and particle positions at a time determine the history of a system, these fine sets are individual possible worlds.

[33] In quantum statistical mechanics one generally works with probability density operators, not probability measures over states, and the density operators underdetermine the probability measures that could be used to justify it. But a satisfactory justification for the density matrix used in quantum statistical mechanics can (and perhaps must) be obtained from a measure over states. For one way to do this, see Tumulka and Zanghi (2005).

[34] See Berndl et al. (1995).

Since the fine sets of quantum statistical mechanics are the coarse sets of Bohmian mechanics, gluing the two theories together is simple. Let the coarse sets of quantum statistical mechanics be the coarse sets of the combined theory, and let the fine sets of Bohmian mechanics be the fine sets of the combined theory. Then we get the appropriate measures for the combined theory, statistical Bohmian mechanics, by essentially taking the product of the quantum statistical mechanical measures and the Bohmian mechanical measures.

We can now sketch how statistical Bohmian mechanics constrains our priors. First we determine our subjective prior in statistical Bohmian mechanics, and divide this subjectively among the coarse sets of the theory. Then we divide our prior in each coarse set among its fine sets in accordance with the chances. If these fine sets are individual possible worlds, we don't need to divide our priors any further; if not, we divide it between the individual worlds subjectively. So the chances of statistical Bohmian mechanics constrain how our priors in sets of possibilities that share the relevant structural properties are divided up among sets of possibilities which also have the same wave function and particle positions.

A similar procedure can be used to obtain the complete chance theory of quantum statistical mechanics and other quantum mechanical interpretations.[35] For genuinely indeterministic interpretations, for example, we can obtain the chances of histories that share the relevant structural properties by essentially taking the product of the quantum statistical mechanical chances for their initial wave functions and the stochastic chances of their histories given those initial

[35] By this I mean *complete* quantum mechanical interpretations, not interpretations whose content hangs on vague terminology or which are otherwise imprecise. I take it that I am under no obligation to provide a precise account of the chances of chance theories which are not themselves precise.

On some quantum mechanical interpretations the status of quantum statistical mechanics changes to the extent that a procedure for gluing quantum mechanics to quantum statistical mechanics isn't needed. For example, Albert (2001) has argued that if we adopt the GRW interpretation of quantum mechanics an additional statistical theory isn't needed to explain 'statistical mechanical' phenomena.

wave functions.

2.6.3 Consequences of M1-M4

To get a better feel for the pros and cons of M1-M4, let's look at some of the consequences of this account.

Chances and their Grounds

Every chance distribution assigns values to propositions. And every chance distribution is grounded by some proposition—TH for Lewis, TB for me. Is there any relation between the values the chance distribution assigns to propositions and the proposition that grounds the distribution?

Lewis thinks there is, and I agree. But Lewis is only able to derive constraints of this kind by making use of the Principal Principle. Lewis's *metaphysical* claims about chance, L1-L6, tell us very little about this relation. Indeed, aside from L5, these principles impose virtually no constraint on which grounds and chance assignments are compatible.

This strikes me as an unhappy state of affairs. While we may be able to deduce some constraints of this kind by looking at the form of a chance-credence principle, this principle isn't the reason why chances satisfy these constraints. A chance-credence principle takes the form it does because of how grounds and chances are related, not the other way around. If grounds and chances are related in certain ways, it would be nice if this followed directly from one's *metaphysical* account of chance.

M4 gives us these constraints directly. It relates the grounds of a chance distribution to the chances the distribution assigns. For example, M4 entails that the grounds of a chance distribution is always assigned a chance of 1 by that distribution:

$$ch_G(G) = \frac{m(G)}{m(G)} = 1. \tag{2.26}$$

M4 also entails that the chances assigned by different distributions are related in a way that naturally mirrors the relations between the propositions that ground them. So for dynamical chance theories of the kind Lewis entertains, M4 entails that the chance distributions that obtain at different times are appropriately related. Letting I_{12} be the history between times t_1 and t_2:

$$ch_{TH_2}(A) = \frac{m(H_2A)}{m(H_2)} = \frac{m(H_1I_{12}A)}{m(H_1I_{12})} = ch_{TH_1}(AI_{12}). \tag{2.27}$$

Likewise, M4 ensures that the Basic Principle is compatible with Bayesianism; i.e., that there aren't chance distributions which, together with BP, requires agents to update their beliefs in non-Bayesian ways. So:

$$\begin{aligned} cr_{TBE}(A) &= ch_{TBE}(A) && (2.28)\\ &= \frac{m(BEA)}{m(BE)} && (2.29)\\ & && (2.30) \end{aligned}$$

and

$$\begin{aligned} cr_{TBE}(A) &= \frac{cr_{TB}(AE)}{cr_{TB}(E)} && (2.31)\\ &= \frac{ch_{TB}(AE)}{ch_{TB}(E)} && (2.32)\\ &= \frac{\frac{m(BEA)}{m(B)}}{\frac{m(BE)}{m(B)}} && (2.33)\\ &= \frac{m(BEA)}{m(BE)}. && (2.34) \end{aligned}$$

Apparent Chance Claims and Claims About Chance

A consequence of the account I've proposed is that some claims which appear to be about particular chance assignments will actually be claims about the kinds of chances a theory assigns. Consider the GRW interpretation of quantum mechanics. The theory appears to make chance assignments of the following form:

"If the wave function of the universe is ψ, then there's an x chance of the wave function being ψ' one minute later." But the coarse sets of GRW are worlds with the same initial wave function, and worlds with different initial wave functions could evolve into the same state ψ at some later time. (For simplicity, I'm leaving implicit everything else that we need to fix to get a coarse set.) But then we can't use "the wave function of the universe is ψ (at some time)" as the second grounding argument, B, because it picks out worlds which aren't contained in a single coarse set. So on my account this chance assignment will be undefined. And if we insist on cashing out chance claims of this sort as particular chance assignments, my account will be unable to accommodate them.

That said, we can make sense of these claims as claims about the chances the theory assigns. When we say "if the wave function of the universe is ψ, then there's an x chance of the wave function being ψ' one minute later", we're noting that given any initial wave function, and given the fact that the wave function is ψ at some time t, then the chance of ψ' at (t+1 minute) according to GRW will always be x.[36]

Note that Lewis's account has similar consequences. On Lewis's account, chances are assigned at a particular time. So chance claims like the one given above, "if the wave function of the universe is ψ, then there's an x chance of the wave function being ψ' one minute later", won't be well-defined, since we haven't picked out a particular time at which ψ holds. But, again, we can understand this to be a claim about the chances that GRW assigns. I.e., for any history up to a time t such that the wave function of the universe at t is ψ, the chance of ψ' a minute later according to GRW will be x.

[36] Note also that you'll get the appropriate chance-credence ties from these kinds of chance claims. I.e., if you know that GRW holds, and you're certain that the wave function at t is ψ, but you're uncertain of what the initial wave function is, it will still follow that your credence that the wave function is ψ' at (t+1 minute) should be x.

Conditional and Unconditional Chance

Another consequence of M1-M4 is that every well-defined conditional chance $ch_{TB}(A|E)$ will correspond to an unconditional chance $ch_{TBE}(A)$ (see Appendix 9.9 for details).

This appears to be at odds with our ordinary picture of chance. Consider a world with dynamical chances. At this world there are only two chance events: a pair of fair coin flips, which occur at different times. Let the chance of any particular outcome of these tosses, relative to the initial conditions, be $\frac{1}{4}$. Given the structure I propose, each of these pairs of outcomes corresponds to a fine set, and these four fine sets will lie within a coarse set. But if what I say is right, there is a well defined chance for any union of these fine sets, A, relative to any other union of these fine sets B. For example, there will be a chance assigned to the double heads (HH) result relative to $B =$ (HH∨HT∨TH).

That might seem strange. There's at least an initial intuition that "the chance of HH relative to (HH∨HT∨TH) is $\frac{1}{3}$" isn't a metaphysically fundamental chance. Rather, it's just a conditional chance we've derived from the real chances—the dynamical chances of how a system may evolve over time.

Put another way, it seems metaphysically possible for there to be a chance theory which assigns a $\frac{1}{4}$ chance to each of the four outcomes relative to (HH∨HT∨TH∨TT), but which doesn't assign an unconditional chance to HH relative to (HH∨HT∨TH). On my account, such a chance theory is impossible.

Contrast this with Lewis's account, according to which such chance theories are not only possible, they're mandatory. On Lewis's account, the chance of HH relative to (HH∨HT∨TH) will *never* be defined, since (HH∨HT∨TH) doesn't correspond to a history up to time.

What should we think about this consequence? It's unclear. The initial intuition is that which chances are 'merely' conditional and which are not is a substantive issue. But there is also some intuitive appeal to dismissing the distinction as metaphysically immaterial.

This issue comes up in the debate regarding the ABL account of quantum mechanics. The ABL theory assigns chances to events using both future and past information. As Aharonov, Bergman and Lebowitz have noted, these chances are identical to conditional chances of the standard time asymmetric accounts. The ABL chance of A relative to past information P, future information F and the other things we need to specify B, will be the same as the standard chance of A conditional on F relative to P and B: $ch_{(ABL)PFB}(A) = ch_{(STD)PB}(A|F)$. Given this, a number of physicists—including, at times, the authors of the theory—have argued that there isn't a genuine difference between ABL and the standard theory.

My account supports this claim: on my account the chances of the ABL theory are identical to the chances of the standard approach. The standard approach provides well-defined chances conditional on future information, and since every conditional chance corresponds to an unconditional chance, it follows that the standard approach assigns unconditional chances relative to future information, just like ABL. Lewis's account, on the other hand, leads to the opposite conclusion: a world where ABL holds is qualitatively different from one where the standard theory holds.

2.6.4 Can M1-M4 Accommodate Ordinary Claims About Chance?

It's generally assumed that an adequate account of chance must accommodate our ordinary chance talk. For instance, it's assumed that any adequate theory of chance should be able to give straightforward truth conditions for the ordinary claim "there's a 50% chance of heads".

On Lewis's account there seems to be a straightforward answer. Since chances can be indexed by worlds and times, we can take the utterence of “there’s a 50% chance of heads” to indexically pick out a world and a time. If the chance of heads picked out by that world and time is $\frac{1}{2}$, then the claim is true. Otherwise, the claim is false.

On my account, things look less straightforward. Since we can’t always index chances to a world and time, it’s less clear how to pick them out. There are some tricks available, however. For example, we might take the utterence to indexically pick out the chance theory that holds at the world of the utterence, and the total evidence of the speaker at the time of utterence, and then take the claim to be true *iff* the conjunction of the chance theory and that total evidence yields a chance of $\frac{1}{2}$ for heads.

But before we waste time working out whether this analysis is adequate, let’s return to the claim we started with: that providing an account of ordinary chance claims is a desideratum of an adequate chance theory. This is only plausible if our ordinary use of the term “chance” refers to the notion Lewis and I are concerned with: *nomological chances*, the things that play a role in physical theories like quantum mechanics. But our ordinary claims about chance aren’t generally referring to nomological chances. Indeed, our ordinary use of the term “chance” is a mess: such a mess that it’s unlikely that either of the analyses offered above—for Lewis’s account or for mine—is going to be adequate. So expecting an account of nomological chance to account for ordinary chance claims is a mistake.

A more comprehensive discussion of these issues can be found elsewhere (see Hawthorne (2007)), but here are some brief remarks. Suppose a weather forecaster says:

> “Make sure to bring an umbrella; there’s a 30% chance of rain.”

This is a reasonable statement to make even if the speaker is convinced that there are no non-trivial nomological chances (such as a Lewisian about chance who thinks the laws are deterministic). And the same is true of claims like:

> "20 years ago, Bobby Fisher played a game of chess with the president. The outcome of the game is unknown, but there is a good chance he won."

> "You should be careful about that cut; there's a good chance you might get tetanus."

At first pass one might think that these are simply epistemic uses of the word "chance", and that in these contexts "there's a good chance that" means something like "I have a high credence that". But this can't be the whole story. Consider:

> "You should be careful about that cut; there's a good chance you might get tetanus." (Reply:) "No there isn't; I've been vaccinated."

If the use of "chance" here is purely epistemic, then the reply makes little sense. The first speaker is just reporting her current degree of confidence in whether the second speaker will get tetanus, and it isn't reasonable to disagree with *that*. Since the reply *is* reasonable, the original statement cannot just be a declaration of the first speaker's credences. Moreover, the first speaker was *wrong*: since the second speaker was vaccinated, there *isn't* a good chance that she'll get tetanus. Again, this reaction makes little sense if the first speaker is just declaring her credences. So we aren't using "chance" in a purely epistemic way in these cases. Nor are we using "chance" to mean the kind of objective chances that Lewis and I are concerned with.

So how are our ordinary chance claims relevant to nomological chances? Claims like "there's a 30% chance of rain" aren't relevant: the notion of chance being employed is not the notion we're interested in. And ordinary claims like "there's a 50% chance of heads" are no different from claims like "there's a 30% chance of rain"—both can be intelligibly made by someone who thinks there are no non-trivial nomological chances, and neither is purely epistemic—someone could make either claim and be wrong. So ordinary claims about the chances of coin flips are no more relevant than claims about the chance of rain.

Lewis and I are interested in the objective chances that play a role in the laws of nature. Ordinary talk about "chance" has little bearing on this. So providing truth conditions for ordinary chance claims like "there's a 50% chance of heads" is not a desideratum for a satisfactory account of nomological chance.

2.7 Chance, Credence and Sleeping Beauty

In the sleeping beauty case it's uncontentious that something like the Principal Principle applies on Sunday, and thus that Beauty should have $\frac{1}{2}/\frac{1}{2}$ credence in heads and tails. Some of the sleeping beauty literature has focused on whether the Principal Principle should also apply after Beauty wakes up on Monday.[37] The question is whether she gets admissible evidence when she wakes up on Monday. If so, the thought goes, the Principal Principle should still apply, and her credence in heads and tails should remain $\frac{1}{2}/\frac{1}{2}$.

In the preceding discussion we've found that a satisfactory chance-credence principle doesn't need an admissibility clause. And, if we desire, we can use such a principle to provide a precise characterization of admissible evidence. So if the literature just mentioned is correct, we should now be able to figure out what our credences should be in the sleeping beauty case.

Unfortunately, matters are not so straightforward. Recall the Basic Principle:

$$\text{BP: } cr_G(A) = ch_G(A), \text{ if } ch_G(A) \text{ is defined} \tag{2.35}$$

On one side we have a credence function; on the other, a chance distribution.

[37] See Lewis (2001) and Dorr (2002).

Now consider the variables that appear in this equation. On the chance side, both the objects and the grounds are restricted to *de dicto* propositions. On the credence side, there is no such restriction: neither the objects of our credence nor our total evidence needs to be a *de dicto* proposition.

So how does BP apply in cases where our total evidence is not a *de dicto* proposition? BP only applies when the variables on both sides match. Since chance distributions are only well-defined when A and G are *de dicto* propositions, BP will only place a direct constraint on our credences when A and G are *de dicto* propositions. If our total evidence doesn't match the grounds of any well-defined chance distribution, then, as we've seen, we use our updating rule to figure out what our credences should be. But now things get tricky: in the cases we considered earlier, we assumed we were updating using standard Bayesian conditionalization. And as we've already seen, that's only viable when we're restricting our attention to *de dicto* beliefs. When we take self-locating beliefs into account, we need to employ a more sophisticated updating rule, a *de se* updating rule. And what beliefs we end up with in situations like the sleeping beauty case will depend on the *de se* updating rule we adopt.[38]

So considerations involving chance won't yield the answer to the sleeping

[38] One might try to understand BP this way instead: take BP to apply iff the minimal set of worlds compatible with our total evidence matches the grounds of some chance distribution. Or, letting '$\overline{e}$' stand for the minimal set of worlds compatible with the *de se* proposition e, then BP should read:

$$\text{BP}^*: \; cr_{\overline{e}}(A) = ch_{\overline{e}}(A), \text{ if } ch_{\overline{e}}(A) \text{ is defined} \tag{2.36}$$

This way of understanding BP yields a definite answer to the sleeping beauty case. Since we don't eliminate any doxastic worlds when we wake up, $\overline{e}_{SUN} = \overline{e}_{MON}$, and thus our credences in heads/tails on Monday will be the same as they were on Sunday: $\frac{1}{2}/\frac{1}{2}$.

If we consider the special case of (2.36) that we get when e is a *de dicto* proposition, we get the version of BP described in the text. So (2.36) is strictly stronger. But the extra strength built into (2.36) isn't needed: with a *de se* updating rule in hand, and the understanding of BP described above, we'll know precisely what our credences should be. (2.36) essentially takes BP and builds in some assumptions about what a good *de se* updating rule should be like. If we like these assumptions, then (2.36) will be plausible but redundant. If we dislike these rules, then (2.36) will be implausible. So regardless of what we think of these assumptions, there is little reason to adopt (2.36).

beauty case. If we want to figure out the dynamics of *de se* beliefs, we'll have to look elsewhere.

Chapter 3
Dutch Books

3.1 Introduction

One way to try to resolve the sleeping beauty case is to assess how Beauty should bet. If her credences lead her to bet in such a way as to incur a sure loss—if her credences leave her vulnerable to a Dutch book—then one might argue that her credences are defective.

In what follows, I'll evaluate the bearing of betting arguments on the sleeping beauty problem, in three rounds. The first round examines the betting arguments offered by Hitchcock (2004) in favor of the thirder and against the halfer. The second round follows Arntzenius (2002), and looks at what Beauty should do from a decision theoretic standpoint. The third round evaluates the conclusion Arntzenius draws from this. I'll end by assessing the implications of the above on the sleeping beauty problem.

3.2 Round 1: Hitchcock

Hitchcock (2004) provides a betting argument in favor of the thirder response to the sleeping beauty problem.

He considers the following betting situation. A bookie undergoes the experiment with Beauty, and offers her a bet on Sunday night, as well as every time they wake up. Hitchcock demonstrates that if (i) Beauty bets in the usual way on Sunday night, and (ii) when she wakes up, Beauty is willing to pay $\$\frac{1}{2}$ for a bet that pays \$1 if heads comes up, then the bookie can construct a Dutch book against her. More generally we can show that if Beauty takes anything other than $\$\frac{1}{3}$ to be a fair price for a \$1 bet when she wakes up she can be Dutch booked (see Appendix 9.10). (By "fair price" I mean the highest value that an agent is willing to buy such a bet for, and the lowest value that she is willing to sell the bet for. I will grant the usual assumption that these amounts are the same.)

Dutch books aside, if we repeat the experiment a number of times, frequency considerations alone indicate that Beauty is likely to lose money if she bets in any other way. I.e., suppose she pays $\$\frac{1}{2}$ for a \$1 bet on H every time she wakes up, as Hitchcock takes the halfer to recommend. Since she's woken up twice every time tails comes up, but only once every time heads comes up, she'll lose twice as often as she wins. And since she loses and wins the same amount, we can expect her to lose money in the long run, by an average of $\$\frac{1}{4}$ per trial.

With these results in hand, Hitchcock assumes that, after waking up on Monday morning, Beauty should accept bets in accordance with her credences in the usual way. That is, if her credence in A is x, she should consider a bet on A which pays \$1 to be worth \$x. He then concludes that unless Beauty's credences in heads and tails are $\frac{1}{3}/\frac{2}{3}$, she'll be susceptible to a diachronic Dutch book.

This argument is valid, but there are immediate worries about whether it is sound. In particular, it's not clear that Beauty ought to bet in accordance with her credences in the standard way when she wakes up on Monday. One might argue that the bets Hitchcock considers double count tails results. Beauty should take the payoff for tails results to be twice as large as the bookie claims, since the tails outcome effectively happens twice. From this perspective, the Dutch book Hitchcock presents will end up telling against the thirder, not the halfer. The halfer will bet in just the way that Hitchcock claims the thirder will, and will escape unscathed. The thirder, on the other hand, will favor tails twice as strongly as Hitchcock suggests he will, and will be vulnerable to a version of

Hitchcock's Dutch book.

Here is one way to spell out this argument. Consider the following five pairs of cases. In each case, we want to know what a rational agent should take fair odds to be.

Let's start with two canonical cases:

> 1A. A bet on a fair coin toss.
>
> 1B. A bet on an unfair coin which lands tails $\frac{2}{3}$ of the time.

In these cases, fair odds in H/T are 1:1 (for 1A) and 1:2 (for 1B), respectively.

Now let's consider a more interesting pair of cases:

> 2A. A bet on a fair coin toss, but with the following twist. A robot automoton has been given access to your bank account. If the coin lands tails, the automoton will bet on the outcome of the coin toss using your bank account, in such a way as to imitate your previous betting behavior. I.e., it will bet on the same outcome you have bet on.
>
> 2B. A bet on an unfair coin which comes up tails $\frac{2}{3}$ of the time, and where an automoton has been given access to your bank account, as described above.

Let's start with 2A. What should you take fair odds in H/T to be? If the coin lands tails, and you've purchased a bet on tails, you'll end up making double the profit you would have made from a tails bet in 1A. On the other hand, if you've bet on heads and the coin lands tails, you'll end up losing twice as much as you would have in 1A, since the automoton will buy a losing bet identical to your own. So, compared to 1A, tails results are twice as valuable as heads results. Since fair betting odds in 1A are 1:1, our betting odds in 2A should be 1:2.

What about 2B? 2B is related to 1B in the same way as 2A is related to 1A. And just as tails results are twice as valuable in 2A as they were in 1A, tails results are twice as valuable in 2B as they were in 1B. Since fair betting odds in 1B are 1:2, fair betting odds in 2B are 1:4.

> 3A. A case like the sleeping beauty case, but where instead of being woken up twice if the coin comes up tails, a duplicate of you will be created. (I'll call this the duplication version of the sleeping beauty case.) And while the duplicate will also be offered a bet on the outcome of the coin toss, she will not bet with your bank account. Finally, assume that you are a halfer.
>
> 3B. A case like (3A), but where you're a thirder, not a halfer.

What are fair odds in H/T in 3A? In this case, the potential existence of the duplicate seems irrelevant. Since the duplicate does not bet with your bank account, how she bets is unimportant. As far as you're concerned, this situation is no different than 1A, where you're simply betting on a fair coin toss. So you should adopt the same odds in 3A as you did in 1A: 1:1.

3B is related to 1B in the same way as 3A is related to 1A. As before, your odds in 3B should be the same as your odds in 1B. So your odds in 3B should be 1:2 in H/T.

> 4A. A case like 3A, but where you and the duplicate are hooked up to robotic automotons of the kind described in case 2A. (So: a Duplication SB case where you and the duplicate don't share bank accounts, where you're a halfer, and where both you and the duplicate are hooked up to robotic automotons).
>
> 4B. A case like 4A, but where you're a thirder, not a halfer.

What are fair odds in H/T in 4A? As in 3A, the existence of the duplicate is irrelevant, since you're not sharing bank accounts. And as in 2A, the existence of the automoton effectively doubles the value of tails outcomes. So your odds in this case should be like your odds in 3A, but with tails weighed twice as heavily. I.e., your odds in H/T should be 1:2.

4B is similar: the existence of the duplicate is irrelevant, but the existence of the automoton effectively doubles the value of tails outcomes. So your odds in this case should be like your odds in 3B, but with tails weighed twice as heavily. Thus your odds in H/T should be 1:4.

> 5A. A case like 3A, but where you and the duplicate share bank accounts. (So: a Duplication SB case where you and the duplicate share bank accounts, and where you're a halfer.)
>
> 5B. A case like 5A, but where you're a thirder, not a halfer.

What odds should you adopt in 5A? In this case you aren't hooked up to a robotic automoton. But the duplicate will behave just like a robotic automoton: if the coin lands heads she'll place a bet on the coin toss identical to the bet you placed, using money from your bank account. So betting-wise, this case is identical to 3A, the case where you're hooked up to an automoton, and you should accept the same odds. Thus your odds in H/T should be 1:2.

5B bears the same relation to 3B as 5A does to 3A. As with 5A and 3A, one's odds in 5B and 3B should be the same. Thus your odds in H/T should be 1:4.

Cases 5A and 5B are essentially identical to the original sleeping beauty case, for the halfer and the thirder, respectively. Your Monday and Tuesday temporal parts will both bet with the same bank account, on the same event. So the odds the halfer and the thirder should accept as fair in the sleeping beauty case are 1:2 and 1:4, respectively—twice the odds that Hitchcock assumes. Thus it seems that Hitchcock's implicit assumption is false: when Beauty wakes up on Monday, she shouldn't bet in accordance with her credences in the standard way. And the Dutch book Hitchcock provides actually tells in favor of the halfer and against the thirder, not the other way around.

3.3 Round 2: Decision Theory

Should we conclude that betting arguments favor the halfer? Not yet. The problem with Hitchcock's argument is that the crucial premise—how your credences are related to how you should bet—relies on a vague assessment of how one ought to bet in certain situations. But the argument by analogy for the halfer I gave relies on the same kinds of assessments. If we want to figure out whether betting odds favor one side or the other, we need to start with something more principled.

In ordinary cases, claims about the relation between credences and fair betting odds are justified by appealing to decision theory. If we take an agent's utility to be linear in dollars, and we are given the agent's credences, then we can deduce the betting odds which make the expected utility of each side of a bet the same. This gives us the agent's fair betting odds. Similarly, we can work out that if her credence in A is x, she should consider a bet on A which pays $1 to be worth $x.

How should Beauty bet according to decision theory? As Arntzenius (2002) has pointed out, it depends on whether you adopt causal or evidential decision theory. Indeed, causal decision theory grounds the key premise in Hitchcock's argument for the thirder, and evidential decision theory grounds the key premise in my argument for the halfer.

In Hitchcock's argument, the key premise was that Beauty should bet in accordance with her credences in the standard way. If we're causal decision theorists, this assumption is justified. Beauty's betting behavior on Monday is causally independent from her betting behavior on Tuesday. The odds she accepts on one day have no causal bearing on the odds she accepts on the other. So when she wakes up on Monday morning, she should accept whatever bets she would normally accept: if her credences in heads and tails are $\frac{1}{3}/\frac{2}{3}$ she should take fair odds to be 1:2, and if her credences are $\frac{1}{2}/\frac{1}{2}$, she should take fair odds to be 1:1. Thus the causal decision theorist will conclude that the Dutch book tells against the halfer. (A derivation of this result is provided in Appendix 9.11.)

If we're evidential decision theorists, the assumption that Beauty should bet in accordance with her credences in the standard way is not justified. Beauty's betting behavior on Monday is excellent evidence for her betting behavior on Tuesday. And when she evaluates how to bet, she'll take into account the fact that her other temporal part will bet in the same way. Given this, she'll take tails outcomes to be twice as valuable as the causal decision theorist. So if her credences in heads and tails are $\frac{1}{3}/\frac{2}{3}$ she should take fair odds to be 1:4, and if her credences are $\frac{1}{2}/\frac{1}{2}$, she should take fair odds to be 1:2. Thus the evidential decision theorist will conclude that the Dutch book tells against the thirder. (A derivation of this result is provided in Appendix 9.11.)

In my argument for the halfer the key assumption was that, with respect to betting, having a bank account-sharing duplicate is just like having a bank account-sharing automoton. And so with respect to betting, case 5A is just like case 2A. But while this assumption is justified if you're an evidential decision theorist, it's not justified if you're a causal decision theorist. The way the automoton bets is causally determined by how you bet, but the way the duplicate bets is not. So for the causal decision theorist, case 5A is not like case 2A; and while you're justified in adopting 1:2 odds in 2A, you're not justified in adopting those odds in 5A.

So, as Arntzenius (2002) notes, the bearing of Dutch books on sleeping beauty hangs on the kind of decision theory one adopts. Evidential decision theorists will conclude that Dutch books can be made against the halfer, and standard causal decision theorists will conclude that Dutch books can be made against the thirder.

Of course, "causal decision theory" is not a single theory: there are a number of different versions of causal decision theory on offer. Normally the differences between these theories aren't attended to, since they all seem to yield more or less the same consequences. But the differences between them might be crucial when evaluating the status of betting arguments in the sleeping beauty case. For instance, while standard causal decision theory seems to support the thirder, some versions of causal decision theory, such as the "counterfactual decision theory" Arntzenius (2002) refers to, might support the halfer.

3.4 Round 3: Arntzenius's Conclusion

Should we conclude that one's position on sleeping beauty depends on what decision theory one adopts? Both Arntzenius (2002) and I think not, but we disagree on the moral to draw. Arntzenius concludes:

> "It seems rather odd that SB's degrees of belief would depend on the decision theory that she accepts. Surely if she changes her mind about which decision theory is correct she should not thereby be forced to change her epistemic state with respect to heads. Surely changing her mind about decision theory does not entail changing her mind as to what the world is like with respect to outcomes of coin tosses. Thus it seems more plausible to say that her epistemic state upon waking up should not include a definite degree of belief in heads."[1]

I agree that it's not plausible to think that if Beauty changes her mind about decision theory, she should change her mind about what the world is like. But the lesson is not that Beauty's credences ought to be indeterminate. Rather, the moral is that betting arguments have little bearing on what our credences ought to be. To see why, let's look at Arntzenius's argument in more detail.

3.4.1 Arntzenius's Argument

We can see Arntzenius's conclusion as an argument by *reductio*, employing six premises. The first two premises are uncontentious. They simply state the results shown in Appendix 9.11.

[1] Arntzenius (2002), p.61

P1. If Beauty adopts evidential decision theory, and she has a precise credence in heads/tails other than $\frac{1}{2}/\frac{1}{2}$, then she is vulnerable to a diachronic Dutch book.

P2. If Beauty adopts causal decision theory, and she has a precise credence in heads/tails other than $\frac{1}{3}/\frac{2}{3}$, then she is vulnerable to a diachronic Dutch book.

The third premise is, I take it, uncontentious as well:

P3. In any situation, there's at least one credence state that is rationally permissible.

The fourth premise expresses Arntzenius's sentiment in the quote given above—that one's decision theory and one's credences in heads and tails should be independent.

P4. Which credences in heads and tails are rationally permissible in the sleeping beauty case shouldn't depend on what kind of decision theory Beauty adopts.

We can't conclude anything from these four premises alone, since P1 and P2 are orthogonal to P3 and P4: P1 and P2 are about bets, P3 and P4 are about rationally permissible credences. To bring to them into contact we need to add a fifth premise:

P5. If having credences of type X makes you vulnerable to a diachronic Dutch book, then having credences of type X is irrational.

These five premises almost bring about a contradiction, but not quite. P1, P2 and P5 entail that if Beauty adopts evidential or causal decision theory, then she's irrational if she has precise credences in heads and tails other than $\frac{1}{2}/\frac{1}{2}$ or $\frac{1}{3}/\frac{2}{3}$, respectively. But if imprecise credences are rationally permissible, then P1, P2, P3, P4 and P5 can all be true simultaneously. To get a contradiction we need to add a sixth premise:

P6. Beauty should have precise credences in heads and tails.

Now we do get a contradiction. Arntzenius takes this to provide a *reductio* against P6, concluding that Beauty shouldn't have precise credences in heads and tails.

3.4.2 Evaluating Arntzenius's Argument

I don't think the rejection of P6 is warranted. Here's why.

Consider the following decision theory, "weird decision theory". If you have precise credences, then weird decision theory's recommendations are identical to those of evidential decision theory. If you have imprecise degrees of belief, then weird decision theory tells you to choose the act with the highest utility outcome. So, for example, given imprecise credences about the outcome of a bet, it recommends that you choose a bet with either a $0 outcome or a $1 outcome over a bet with either a $0.99 outcome and a $0.98 outcome (assuming utility is linear with dollars.)

Now, the following is a consequence of weird decision theory:

P7. If Beauty adopts weird decision theory, and she has a imprecise credence in heads/tails, then she is vulnerable to a diachronic Dutch book.

Here's an example of such a Dutch book. On Monday morning a bettor offers you a ($0.01 if heads and -$1 if tails) bet, and then a (-$1 if heads and -$0.01 if tails) bet. If you have imprecise credences in the outcome of the coin toss, weird decision theory will tell you to accept both, guaranteeing a loss of $0.99.[2]

[2] Does this bet need to be diachronic? It depends. You won't accept the two bets just given if offered as a single bet (and assuming that the other option is to take no bet at all). And if

P1-5 and P7 yield a contradiction without P6. P5 and P7 entail that according to weird decision theory, Beauty is irrational if she has to imprecise credences in heads and tails. And as we've just seen, P1-P5 entail that Beauty is irrational if she has precise credences in heads and tails. Since P1-5 and P7 yield a contradiction, and P7 is plainly true, we have to discard one of P1-P5. And if we have to discard one of P1-P5 anyway, then the *reductio* argument offered by Arntzenius against P6 is not sound.

Now, which one of P1-P5 should we throw out? P1 and P2, like P6, are uncontentious. I take P3 to be uncontentious as well. That leaves us with either P4 or P5. Arntzenius explicitly endorses P4, and I'm inclined to agree with him. So this leaves P5: "if having credences of type X makes you vulnerable to a diachronic Dutch book, then having credences of type X is irrational". We should reject P5.

This premise should have struck us as suspicious from the start. First, our reasons for adopting P4 tell directly against P5. What seems implausible about the claimed relation between Beauty's beliefs about decision theory and what Beauty ought to believe about the coin tosses is that they seem to be about entirely different things. One makes claims about how Beauty ought to act as a *prudentially* rational agent, the other is a claim about what Beauty ought to believe as an *epistemically* rational agent. And, as many people have noted, prudential rationality is orthogonal to epistemic rationality. If being confident that I'll win the race makes me perform better, then I have a prudential reason to be confident that I'll win the race. But I don't have an epistemic reason to be confident I'll win the race: just because it's in my interest to believe it doesn't mean it's more likely to be true.

we take simultaneous bets to be evaluated together, which is natural, then this won't yield a synchronic Dutch book. Like the other Dutch books considered above, it will only work as a diachronic Dutch book.

But if this is why we're inclined to adopt P4, then it's hardly plausible to claim that vulnerability to Dutch books gives us a reason to adopt certain credences. Vulnerability to Dutch books is, if anything, a matter of prudential rationality, while what our credences ought to be (in the sense we're concerned with) is a matter of epistemic rationality. If we like P4, we should never have accepted P5 in the first place.

Second, and more importantly, we have a number of independent reasons to reject P5. There's a large literature of articles pointing out the defects in diachronic Dutch book arguments. This tells against P5, since P5 essentially claims that these diachronic Dutch book arguments are sound.[3]

The right moral to draw from Arntzenius's observations, I think, is one that's been made before in other contexts: being disposed to lose money while gambling has little to do with whether one is an epistemically responsible agent. Vulnerability to Dutch books may give us reason to doubt that we're prudentially optimal, but it gives us little reason to doubt that we're epistemically rational.

3.5 Assessing Dutch Books

Dutch book arguments fail to resolve the sleeping beauty problem.

Betting arguments such as the argument offered by Hitchcock (2004) and the argument by analogy I provided both implicitly rely on premises to which they're not entitled. To get a definitive answer to how one should bet in the sleeping beauty case, one needs to appeal to decision theory. Arntzenius (2002) has shown that which response to the sleeping beauty problem is Dutch bookable depends on

[3] This is a contentious issue, of course, and there are stock replies to these kinds of worries: Dutch books are claimed to be symptoms of epistemic incoherence, and so on. This is not the place to review these responses and their merits. I'll simply state that every defense of Dutch books (diachronic or synchronic) that I have encountered has either begged the question or relied on premises which are plainly false. A critical discussion of diachronic Dutch books can be found in Christensen (1991). A compelling critical commentary on synchronic Dutch books can be found in Weisberg (2006).

what decision theory one adopts. Given standard causal decision theory, Dutch books can be made against the halfer; given evidential decision theory, Dutch books can be made against the thirder.

It's natural to conclude from this that the version of decision theory one adopts determines what one's credences in sleeping beauty cases should be. Arntzenius correctly argues that this conclusion is implausible. Instead, he argues that our credences in the sleeping beauty case should be imprecise. But this argument is untenable, and repairing it leads to a different conclusion: that Dutch books aren't relevant to what our credences ought to be.

So the conventional wisdom about Dutch books turns out to be correct: Dutch books tell us little about what we ought to believe. To figure out what our credences should be in the sleeping beauty case, we'll have to look elsewhere.

Chapter 4
The Framework

In the last two chapters, we saw that hopes of settling the sleeping beauty case without looking at detailed accounts of *de se* beliefs are misplaced. If we want to make progress, we'll have to get our hands dirty.

In this chapter, I'll sketch a formal framework in which to lay out and evaluate the accounts we'll look at. In the first section I'll provide some useful terminology. In the second section I'll spell out some assumptions common to all of the approaches I'll examine. In the third and fourth sections I'll lay out the metaphysical and epistemic frameworks needed for our investigations. In the fifth section I'll present some concepts that will be useful in the discussion to come.

4.1 Terminology

1. Centered worlds are possible worlds paired with times and individuals. The term "individual" is used in Lewis's sense: any possible object counts as an individual.[1] Some centered worlds are centered on rocks, while others are centered on subjects with mental lives and beliefs. Call belief-having centered worlds *epistemic subjects*.

2. Lewis provides two different characterizations of *doxastic worlds* and *doxastic alternatives*. On his first characterization, *doxastic worlds* are the worlds a subject believes might be hers, and *doxastic alternatives* are the centered worlds

[1] Though not every object period. On Lewis's account there are also "impossible" objects, objects which are parts of more than one world. See Lewis (1986*a*), p. 211.

a subject believes might be hers. On his second characterization, *doxastic worlds* are the worlds in which a subject has a non-zero credence, and *doxastic alternatives* are the centered worlds in which a subject has a non-zero credence.

For Lewis these two characterizations are equivalent because he employs non-standard probabilities to represent credences. Lewis suggests that we allow infinitesimal credences, and that we assign an infinitesimal credence to any possibility we don't have a finite credence in, but think might obtain.[2] So on Lewis's account, having a positive credence in a possibility and believing a possibility might obtain are equivalent.

But these two characterizations are not equivalent if we use standard probabilities to represent credences. On the standard approach, having a positive credence in a possibility is a sufficient but not necessary condition for believing the possibility might obtain: a subject can have a 0 credence in a possibility, and still believe that it might obtain. A classic example of this is a rational agent's credence in a countably infinite number of coin tosses all landing heads. If we use standard probabilities to represent credences, her credence in the outcome will be 0, even though she thinks this possibility could obtain.

I will not take a stand here on whether we should employ non-standard probabilities to represent credences. But if we don't employ them, we're left with a dilemma: which of the two characterizations of doxastic worlds and doxastic alternatives should we employ? For the purposes of this work, it will be more convenient to adopt the second characterization. So I will take the *doxastic worlds* of a subject with credences *cr* to be:

$$DW(cr) = \{W | cr(W) > 0\}, \tag{4.1}$$

[2] See citelewis:1986a, for example.

and her *doxastic alternatives* to be:

$$da(cr) = \{c|cr(c) > 0\}, \tag{4.2}$$

where W and c range over worlds and centered worlds, respectively.

(In one place (Appendix 9.13) it will be useful to use the first characterization instead. In that context, I'll add an asterix to the terms to denote this. So a subject's *doxastic worlds** are the set of worlds she believes might be hers. Likewise, her *doxastic alternatives** are the centered worlds she believes might be hers.)

3. Every set of centered worlds corresponds to a *centered* or *de se proposition*. Some of these sets correspond to unions of worlds—are such that, for any world W, either all of the centered worlds located at W are in the set or none of them are. Each of these sets corresponds to a *de dicto* proposition. So every *de dicto* proposition is a *de se* proposition, but not vice versa.

I'll represent centered worlds and *de se* propositions with lower case letters, and worlds and *de dicto* propositions with capital letters. In the chapters to come I'll also employ the following covention: given a centered proposition c, I'll let $\overline{c}$ stand for the (minimal) set of worlds containing c.

4. Consider a sum of the form $\sum_i \frac{a}{b_i+c}$. This sum will not be well defined if there are b_i's such that $b_i + c = 0$. In some cases, we will want to consider sums which simply ignore undefined terms like these. In these cases, I'll indicate that undefined terms are to be ignored by adding a "def" superscript to the summation sign. For example, in the above case:

$$\sum_i^{\text{def}} \frac{a}{b_i + c}. \tag{4.3}$$

4.2 Common Assumptions

All of the accounts we will look at share some common assumptions. Like standard Bayesianism, they apply to idealized agents. Such agents satisfy the following three conditions:

A1. The agent's epistemic state at a time can be represented by a probability function over a space of possibilities. These values, called *credences* or *degrees of belief*, indicate the subject's confidence that the possibility is true, where greater values indicate greater confidence.

A2. The agent's evidential state at a time can be represented by a set of possibilities. The possibilities in the set are the possibilities compatible with the evidence the agent has.

A3*. The space of possibilities in question is the space of centered worlds.

The first two conditions are identical to those assumed by Bayesianism, and described in chapter 1. The third condition is different, however: instead of taking the space of possibilities to be worlds, it takes the space of possibilities to be centered worlds.

These accounts also presuppose the following idealized picture of evidence:

A4. Every epistemic subject has a subjective state.

A5. An epistemic subject's evidence is the set of possibilities compatible with her subjective state. Taking possibilities to be centered worlds, the centered proposition representing a subject's evidence is the set of centered worlds (individuals at a time in a world) that share her subjective state.[3]

[3] In the next section we'll characterize a "same subjective state as" relation that supports these claims. In particular (as required by A4) subjective states are mutually exclusive. Since a subject's evidence consists of the centered worlds compatible with her subjective state, it follows that if a pair of subjects satisfy ($a \notin e \Rightarrow cr_e(a) = 0$) and they have the same credences, then they must be in the same subjective state.

A6. The relevant notion of "subjective state" is an internalist one: an epistemic subject's subjective state is determined by her intrinsic properties.

These claims are controversial. Weatherson (2005) argues that a number of philosophical stances appear to be incompatible with this kind of picture—including externalist accounts of evidential experience, externalist accounts of evidential justification, and some accounts of phenomenal vagueness. So this account of evidence is open to the criticism that it conflicts with a number of philosophical accounts in epistemology and philosophy of mind.

Although I won't defend this picture of evidence in detail, let me note why I don't find this criticism compelling. First, some of the conflicts described above are only apparent. For example, this account can be compatible with externalist accounts of experience. All it needs is for there to be something that loosely matches our intuitive notion of a "subjective state" that can be treated in an internalist manner. It doesn't matter whether experiences generally supervene on the intrinsic features of the subjects experiencing them, nor does it matter whether subjective states count as "experiences". Likewise, this account can be compatible with externalist accounts of justification. It requires that subjective states provide a normative constraint on one's credences, but this is compatible with there also being externalist constraints on one's credences, with there being no internalist constraints on knowledge (as opposed to credences), and so on.

Second, this criticism seems somewhat misplaced given the idealizations we're working with. Standard externalist accounts of experience, for example, rely on accounts of content more fine-grained than sets of centered worlds. While such accounts of experience cannot be accommodated by the model we're working with, I don't take this to be a good reason to abandon this approach (see the discussion in section 1.3). Likewise, the accounts we'll look at are ill-equipped to accommodate some standard treatments of vague phenomena. But while expanding these accounts to incorporate one's favorite account of vagueness is an interesting and worthwhile project, it's not the project we're engaged in here.

4.3 Metaphysical Framework

The accounts of *de se* beliefs we'll look at make use of various bits of metaphysical and epistemological structure. In this section I'll describe the metaphysical structure; in the following section, I'll describe the epistemic structure.

We can characterize the space of centered worlds in the following way:

CW1. Let there exist a set of elements, Ω. Call these elements *centered worlds.*

CW2. Let there be a transitive, reflexive and symmetric relation $W(\cdot,\cdot)$ defined over the elements of Ω. Call W the *same world as* relation, and call the sets of elements closed under this relation *worlds*, W_i. These worlds form a partition of Ω.

CW3. Let there be a transitive, reflexive and symmetric relation $T(\cdot,\cdot)$ defined over the elements of Ω. Call T the *same time as* relation, and call the sets of elements closed under this relation *times*, t_i. These times form a partition of Ω.

CW4. Let there be a transitive, reflexive and symmetric relation $I(\cdot,\cdot)$ defined over the elements of Ω. Call I the *same individual as* relation, and call the sets of elements closed under this relation *individuals*, i_i. These individuals form a partition of Ω.

CW5. These three relations uniquely pick out every element of Ω:

$$W(c,c') \wedge T(c,c') \wedge I(c,c') \Leftrightarrow c = c'. \tag{4.4}$$

These three relations (W, T and I) and (4.4) provide Ω with the structure it needs to be a space of centered worlds. It follows from (4.4) that we can think of centered worlds as ordered triples of worlds, times and individuals. Note, however, that the space of centered worlds does not include *every* ordered triple of a world, time and individual. Some times won't exist at some worlds, and some possible individuals won't exist during some times at some worlds. Rather, we should think of the space of centered worlds as every ordered triple consisting of a world, a time that exists at that world, and a possible object that exists at that world during that time.[4]

Many of the details of this characterization stem from the desire to remain neutral with respect to whether or not there are temporal parts and whether counterpart theory or cross-world identity is correct—i.e., whether or not there are 'modal parts'. Insofar as our concern is characterizing the content of *de se* beliefs in terms of centered worlds, we need to make sure that the centered worlds we're dealing with are fine grained enough. If, like Lewis, one adopts counterpart theory and temporal parts, then the possible individuals themselves are fine grained enough to characterize *de se* beliefs, and we can take centered worlds to correspond to possible individuals.[5] We can add world and time indices, as we have above, but they don't add to our expressive power. But if we deny temporal and modal parts this is no longer the case. We need to add time and world indices in order to characterize appropriately fine grained beliefs.

Next, we need a notion of subjective state such that a subject's evidence is the set of centered worlds compatible with her subjective state.

SS. Let there be a transitive and symmetric (but not reflexive)[6] relation $SS(\cdot, \cdot)$

[4] This is the implicit reason for the "individual at that world" caveat in the centered world characterization Lewis (1979) gives in section 10. (He leaves out the "during that time" clause because he's assuming temporal parts.)

[5] Again, I'm understanding "possible individuals" in Lewis's sense; i.e., as any thing which is a part of one and only one world. (See Lewis (1986*a*), p.211.)

[6] SS isn't reflexive because some centered worlds might have no subjective state at all. If so, then $SS(c, c)$ won't always hold.

defined over the epistemic subjects in Ω. Call SS the *same subjective state as* relation. The sets of elements closed under this relation correspond to the different kinds of *evidence*, e_i. While different kinds of evidence are mutually exclusive, they needn't be exhaustive. Thus the different kinds of evidence needn't form a partition of Ω.

The different kinds of evidence won't form a partition of Ω because there are centered worlds without subjective states. (For example, the centered world picked out by this world, this time and my shoe.)

Finally, we need to characterize the continuity relation that epistemic subjects bear to their past selves.

$\mathbf{C}_t$. Let there be a transitive ordered relation $C_t(\cdot,\cdot)$. Call C_t the *temporal continuant of* relation. These relations are asymmetric ($C_i(a,b) \Rightarrow \neg C_i(b,a)$) and hold only between members of Ω that are located at the same world ($C_i(a,b) \Rightarrow W(a,b)$).

C_t is closely tied to our folk notion of personal identity over time. Let $PI(\cdot,\cdot)$ be the relation which holds between any pair of epistemic subjects who are the same person. Then:

$$C_t(a,b) \vee C_t(b,a) \Leftrightarrow PI(a,b). \tag{4.5}$$

Since we'll be looking at several cases involving duplication, fission and fusion, it will be convenient to make some assumptions about how they relate to temporal continuity. For definiteness, I'll assume that temporal continuity persists through fission and fusion, but does not hold between you and any future duplicates. So for example, if you will fission into two "fissiles" a minute from now, they will be temporal continuants of yours. But if a duplicate of you is created a minute from now, it will not be a temporal continuant of yours.

4.4 Epistemic Framework

Now let's turn to the epistemic framework we'll need.

First, we need an algebra over Ω in order to characterize the probability functions that represent credences and hypothetical priors.

EF1. Let there be a σ-algebra $\mathcal{A}$ defined over Ω which contains only measurable sets.

EF2. The credences and hypothetical priors of condition-satisfying subjects are probability measures defined over the elements of $\mathcal{A}$.

Some of the accounts we'll look at also make use of a privileged series of measures:

EF3. Let there be a number of probability measures m_i defined over the elements of $\mathcal{A}$, with one measure m_i for each world W_i. Call m_i the *canonical self-locating measure* over W_i. These measures are such that (i) m_i only assigns positive values to members of $\mathcal{A}$ in W_i, and (ii) if W_i has only a finite number of centered worlds, then m_i is identical to the counting measure over centered worlds.

4.5 Some Useful Concepts

Using the framework we've built up in the past two sections, we can now spell out some useful concepts.

First, the notion of a *temporal successor*. Intuitively, your temporal successors are the subset of your temporal continuants that are in the subjective state you'll be in next.[7] Formally, we can characterize your temporal successors as the

[7] Why "intuitively"? Because strictly speaking, this description isn't right. If your next three subjective states are a, then b and then a again, only the temporal continuants in subjective state a the first time will be temporal successors.

epistemic subjects you bear the *temporal successor of* relation S_t to, and define S_t as:

$$\begin{aligned} S_t(a,b) \quad \textit{iff} \quad & C_t(a,b) \wedge \neg SS(a,b) \wedge \neg\exists c\Big(C_t(a,c) \wedge C_t(c,b) \\ & \wedge\neg SS(c,a) \wedge \neg SS(c,b)\Big). \end{aligned} \tag{4.6}$$

In words: b is a temporal successor of a *iff* b is a temporal continuant of a in a different subjective state, and there's no temporal continuant c in between a and b that's in a different subjective state from either of them.

(4.6) uses C_t to define S_t, but we could also do this the other way around. Given a notion of temporal successors, we can characterize your temporal continuants as the union of your temporal successor, and the temporal successor of your temporal successor, and so on. Or, formally speaking, we can define C_t in terms of S_t as:

$$C_t(a,b) \quad \textit{iff} \quad S_t(a,b) \vee (C_t(a,c) \wedge C_t(c,b)). \tag{4.7}$$

Next, let's characterize the notion of an *n-step temporal successor.* Your temporal successor has a temporal successor: call this your *2-step temporal successor.* And your 2-step temporal successor has a temporal successor: call this your *3-step temporal successor.* And so on.

Following suit, call the evidence that a subject will get next—the evidence that her temporal successor will get—her *1-step evidence.* Likewise, call the ordered pair of the evidence that her successor will get and the evidence that her 2-step successor will get her *2-step evidence*: the next two pieces of evidence she'll get. And call the ordered n-tuple of the next n pieces of evidence she'll get, her *n-step evidence.*

Lewis employs the term "doxastic" to indicate the worlds and centered worlds that a subject has a positive credence in. Or, restricting ourselves to possibilities in which the subject has a positive credence, the possibilities she thinks might be hers. We can also use this terminology with respect to notions we've just introduced. So a subject's *doxastic temporal successors* are the centered worlds that she thinks might be her temporal successors; i.e., the temporal successors of her doxastic alternatives. A subject's *doxastic n-step temporal successors* are the centered worlds that she thinks might be her n-step temporal successors; i.e., the n-step temporal successors of her doxastic alternatives. A subject's *doxastic temporal continuants* are the centered worlds that she thinks might be her temporal continuants; i.e., the temporal continuants of her doxastic alternatives. And a subject's *doxastic n-step evidence* consists of all of the ordered n-tuples that she thinks might comprise her n-step evidence; i.e., the n-step evidence of each of her doxastic alternatives.

It will be useful to have a concise way of referring to the doxastic temporal successors of a subject, so let "$dts(cr)$" be the set of doxastic temporal successors of a subject with credences cr. I.e., letting a and b range over centered worlds,

$$dts(cr) = \Big\{a \;\Big|\; \exists b\Big(b \in da(cr) \wedge S_t(b,a)\Big)\Big\}. \tag{4.8}$$

Finally, we need to characerize a subject's *doxastic epistemic successors*. A subject's doxastic epistemic successors consists of her doxastic temporal successors and any centered worlds at her doxastic worlds that are in the same subjective state as any of her doxastic temporal successors. In the original sleeping beauty case, for example, Beauty's doxastic temporal successors on Sunday night will consist of her temporal continuants right after she wakes up on Monday morning. Beauty's doxastic epistemic successors, on the other hand, consists of her Monday morning temporal continuants *and* her Tuesday morning continuants at the tails worlds, since these centered worlds are located at her doxastic worlds, and are in the same subjective state as her Monday morning continuants.

To refer to a subject's doxastic epistemic successors concisely, let "$des(cr)$" be the set of doxastic temporal successors of a subject with credences cr. I.e., letting a and b range over centered worlds,

$$\begin{aligned} des(cr) \;=\; & \Big\{a \;\Big|\; \big(a \in dts(cr)\big) \vee \Big(\exists a\big((a \in dts(cr)) \\ & \wedge(b \in DW(cr)) \wedge SS(a,b)\big)\Big)\Big\}. \end{aligned} \tag{4.9}$$

A note about the following chapters before we proceed. For ease of exposition, I'll generally discuss the accounts we'll look at in finitary terms. In most cases, the method of extending this discussion to the infinite case is straightforward (replace sums with integrals, etc.).

Chapter 5
Hypothetical Prior Accounts

5.1 The Problem

The standard Bayesian account does not apply to *de se* beliefs. First, the propositions it applies to are sets of possible worlds, and these kinds of propositions aren't fine grained enough to capture self-locating beliefs. But this isn't the only problem. If it were, we could simply replace worlds with centered worlds—replace A3 with A3*—and be done with it. But as we saw in section 1.1, there's a second problem as well: the standard Bayesian account makes certainties permanent, while an adequate account of *de se* beliefs will not.

Here's another way of posing the second problem. We can describe the classical Bayesian updating rule in the following way. When you get new evidence, you eliminate the doxastic worlds incompatible with that evidence, and renormalize your credences in the survivors—assign them values such that the ratios between their credences stay the same, and they add up to 1. This kind of updating only allows for the elimination of possibilities. When you get new evidence you eliminate doxastic worlds, but you never add them. (This is why certainties are permanent: if being certain that P entails that all of your current doxastic worlds are compatible with P, and if you only lose doxastic worlds when you update, then all of your future doxastic worlds will be compatible with P as well.[1]) And

[1] Or virtual certainties, anyway. If your credence in a proposition is 1, then it will be 1 permanently. Whether a credence of 1 entails certainty is a contentious issue, as we've seen in section 4.1.

this feature of the rule remains unchanged if we replace worlds with centered worlds.

But when we take self-locating beliefs into consideration, we want to both eliminate and add possibilities. Say you have an accurate clock in front of you that reads 9 am. Right now, you're certain that it's 9 am, not (say) 9:01 am, or any other time. So all of your doxastic alternatives are located at 9 am; none of them are located at 9:01 am. But a minute later, the clock reads 9:01 am, and intuitively you should now be certain that it's 9:01 am, not 9 am. That is, intuitively, all of your doxastic alternatives should be located at 9:01 am, not 9 am. But this requires your new evidence—that the clock now reads 9:01 am—to not only eliminate your old 9 am alternatives, but to add new 9:01 am alternatives. And the Bayesian "eliminate and renormalize" rule doesn't allow this kind of belief change, since it doesn't allow evidence to add possibilities.

So what we need is a rule that allows both the elimination and the addition of possibilities. One promising route is suggested by the modern formulation of Bayesianism. While the classical characterization of conditionalization (1.1) directly entails that possibilities can only be eliminated, the modern characterization of conditionalization (1.2) does not. On the modern characterization of conditionalization you generate your credences from your hypothetical priors and your evidence. You take your hypothetical priors, set the value of every world incompatible with your evidence to 0, normalize these values, and set your credences equal to the result. This allows for the addition of possibilities: if your current evidence is compatible with worlds that you have 0 credence in, then you gain doxastic worlds when you update. To rule out the addition of possibilities, (1.2) requires the addition of the second clause we saw in section 1.2.

This suggests a natural way of extending Bayesianism to allow for the addition of possibilities. If we adopt the modern characterization of conditionalization but don't adopt the second clause, then we have a Bayesian-style rule which allows for both the addition and the elimination of possibilities. Call this *hp-conditionalization*.

Hp-Conditionalization: A condition-satisfying agent with evidence E and hypothetical priors hp should have the following credences:

$$cr_E(\cdot) = hp(\cdot|E). \tag{5.1}$$

5.2 Two *De Se* Updating Rules: CeC and CoC

Standard conditionalization can't accommodate *de se* beliefs because the space of possibilities it works with is the space of worlds, and it doesn't allow for the addition of possibilities. We can resolve the first problem by replacing the space of worlds with the space of centered worlds, and we can resolve the second problem by replacing standard conditionalization with hp-conditionalization. Together, these responses give us a way to extend standard conditionalization to accommodate *de se* beliefs: we can replace standard conditionalization with hp-conditionalization, and then replace worlds with centered worlds. Call this version of *de se* conditionalization *centered conditionalization*:

Centered Conditionalization (CeC): If a condition-satisfying[2] agent with credences gets evidence e, then her new credence in a centered proposition a, $cr_e(a)$, should be:

$$cr_e(a) = hp(a|e). \tag{5.2}$$

On centered conditionalization you generate your current credences from your hypothetical priors and your current evidence. To get your new credences you take your hypothetical priors in centered worlds, set the credence in every centered

[2] I.e., and agent who satisfies A1, A2, A3*, A4, A5 and A6

world incompatible with your evidence to 0, and then normalize the credences in the remaining doxastic alternatives; i.e., adjust the values such that they sum to 1, and such that the ratios between them are the same as the ratios between your hypothetical priors.

Centered conditionalization, or CeC, is one way to extend hp-conditionalization in order to account for *de se* beliefs. However, CeC and hp-conditionalization are incompatible.[3] To see this, consider a subject with just two doxastic worlds, A and B, with two doxastic alternatives at each world. Assume that her credences are divided equally among alternatives, so that her credence in each alternative is $\frac{1}{4}$ and her credence in each world is $\frac{1}{2}$. What should her credences in worlds A and B be if one of her alternatives at A is eliminated? According to hp-conditionalization her credences in A and B should remain $\frac{1}{2}/\frac{1}{2}$. Her evidence hasn't eliminated any doxastic worlds, so hp-conditionalization will assign the same credences. According to CeC, on the other hand, her credences in A and B should change. After the alternative at A is eliminated, CeC redistributes this credence among alternatives, so that her credence in each alternative is $\frac{1}{3}$. Since she has one alternative at A and two alternatives at B, her credence in A should now be $\frac{1}{3}$ and her credence in B should now be $\frac{2}{3}$.

There is another way to modify hp-conditionalization in order to accommodate *de se* beliefs that avoids this conflict. I'll call it *compartmentalized conditionalization*:

Compartmentalized Conditionalization (CoC): If a condition-satisfying[4] agent gets evidence e, then her new credence in an arbitrary centered proposition

[3] I'm playing a bit fast and loose here. To compare the two rules, we need a way of applying them to the same cases, and this is a bit tricky since they operate over different possibility spaces. To make the comparison I'm assuming adopting the following 'port' of hp-conditionalization in the context of centered worlds: $cr_e(\overline{a}) = hp(\overline{a}|\overline{e})$.

[4] I.e., an agent who satisfies A1, A2, A3*, A4, A5 and A6

a, $cr_e(a)$, should be:

$$cr_e(a) \overset{\text{def}}{=} \sum_{(i|c_i \in a)} hp(\overline{c_i}|\overline{e}) \cdot hp(c_i|\overline{c_i}e). \tag{5.3}$$

On compartmentalized conditionalization, your credences are determined by your priors and your evidence. Compartmentalized conditionalization essentially uses hp-conditionalization to assign credences to worlds, and then divides the credence assigned to each world among its centered worlds in proportion to the centered world's priors.[5]

Here's another way to describe compartmentalized conditionalization, or CoC. Given your priors and current evidence, CoC tells you to determine your new credences in three steps. First, take your hypothetical priors, and set the credence in every centered world incompatible with your current evidence to 0. Second, normalize the credences in the remaining doxastic worlds; i.e., adjust the values assigned to each doxastic world such that they sum to 1, and such that the ratios between them are the same as the ratios between their priors. Finally, normalize your credences in the remaining doxastic alternatives at each world; i.e., at each world adjust the values assigned to the alternatives so that they sum to the credence assigned to that world, and such that the ratios between them are the same as the ratios between their priors.[6]

[5] Recall that your priors, like any probability function, are additive, so your prior in a world is the sum of your priors in the centered worlds at that world.

[6] In an intermediate draft of Meacham (2006), in an attempt to make things easier to follow, I replaced this rule with a simpler rule which divides the credence assigned to a world equally among the remaining doxastic alternatives at that world, instead of in proportion to their priors. But this rule is incompatible with standard characterizations of hypothetical priors. I.e., if *hp* is your priors and you have no evidence ($e = \Omega$), then this rule will not yield *hp* as your credences again.

5.3 Continuity

5.3.1 Continuity and the Passage of Time

De se beliefs raise questions about belief continuity which don't arise in *de dicto* contexts. Consider again the case presented in the introduction, where a subject is watching a clock she knows to be accurate. When the clock changes from 9 am to 9:01 am, the subject discards all of her alternatives at which it's 9 am and replaces them with alternatives at which it's 9:01 am. It seems that her credence in these new alternatives should bear some relation to her credence in the alternatives they've just replaced. But nothing we've said so far requires that this be the case.

Suppose, for example, that the subject watching the clock has only two doxastic worlds, A and B, and that she has only one doxastic alternative at each world. Further suppose that she updates her beliefs using CeC, and that at 9 am her priors in her two alternatives (A(9:00) and B(9:00)) are equal, so her credences in A(9:00) and B(9:00) are $\frac{1}{2}/\frac{1}{2}$. When she sees the clock register 9:01 am, what should her credences in A(9:01) and B(9:01) be? It seems they should be $\frac{1}{2}/\frac{1}{2}$. But there is no reason they have to be this way. Although her priors in A(9:00) and B(9:00) are equal, at 9:01 am these are no longer her alternatives. Her alternatives are now B(9:01) and B(9:01), and nothing we've said so far forces her to have equal priors in these alternatives.

For subjects like us, who have a sense of time passing, every belief change will include a time changing component. As we notice time pass, we replace our old alternatives with new ones located at a later time. Since every evidential change brings an awareness that time has passed, every belief change involves the replacement of old alternatives with new ones. Nothing we've said so far entails that the beliefs of such subjects will be in any way constant—that their credences won't fluctuate wildly simply due to the passage of time. But we think that there should be such constraints; constraints which require a rational subject's beliefs to be diachronically coordinated in the appropriate way. Call constraints of this kind *Continuity Principles*.

A Continuity Principle will take the following form: a subject's credences in her alternatives before and after a belief change ought to be diachronically coordinated when those alternatives are suitably related. For convenience, let us say that an old and new alternative which are suitably related are *continuous* with one another.

To obtain a specific Continuity Principle we need to answer two questions. First, under what conditions are a pair of alternatives continuous? Second, given that a pair of alternatives are continuous, how should our credences in them be correlated?

Let's start with the first question: under what conditions are a pair of alternatives continuous? We seen two candidates for these conditions already: the temporal continuant and successor relations. Given one of these relations, we could hold that a pair of alternatives are continuous *iff* that relation holds between them. I won't take a stand here on what the criteria for continuity in these cases should be. Instead, I'll allow for a variety of Continuity Principles, differing in the standard of continuity they employ. When we come to an argument that requires a Continuity Principle of some kind, I'll provide explicit standards of continuity that are sufficient for these arguments to go through.

Let's turn to the second question: given that a pair of alternatives are continuous, how should our credences in them be related? Consider again the case of the subject watching a clock. In this case we're naturally inclined to assume that her A(9:00) and B(9:00) alternatives are continuous with her A(9:01) and B(9:01) alternatives, respectively. It seems as if her credences in the new alternatives should be the same as her credence in the earlier alternatives they're continuous with. So if her credences in A(9:00) and B(9:00) are $\frac{1}{2}/\frac{1}{2}$, her credences in A(9:01) and B(9:01) should be $\frac{1}{2}/\frac{1}{2}$ as well.

Of course, we don't want to require that credences in continuous alternatives always be the same. Suppose that at 9:01 am the subject learns ¬B, and so has only one alternative at 9:01 am, A(9:01). A(9:01) is continuous with A(9:00), but her credence in A(9:01) should be 1, not $\frac{1}{2}$. So we don't want continuous alternatives to always be assigned the same credences, just to be assigned the same credences when they're in similar evidential situations.

We can capture this intuition by requiring continuous alternatives to have the same priors. Both CeC and CoC are hypothetical prior rules; given one's evidence, they assign credences to alternatives in a manner determined by their priors. So we can get continuous alternatives to have appropriately coordinated credences by requiring them to have the same priors.

But this turns out to be a stronger constraint on priors than we need. We can get the same constraint on credences with a strictly weaker constraint on priors. Let's see how to do this for the two rules we're concerned with.

On CeC, a subject's credences are distributed among her doxastic alternatives in proportion to her priors in those alternatives. Thus the amount of credence assigned to an alternative isn't sensitive to the absolute magnitude of the alternative's prior, only to the *ratio* between its prior and the priors of the other alternatives. So all we need to keep track of is the ratios of the priors between alternatives. Thus on CeC, the Continuity Principle just requires that the ratio of priors between new alternatives be the same as the ratio of priors between any old alternatives they're continuous with.[7]

[7] It may be helpful to see an example of how one could go about imposing this constraint on priors. Let's say that the relevant standard of continuity is the one provided by the temporal successor of relation, S_t. Then for any two centered worlds a and b that are in the same subjective state, the Continuity Principle requires that the ratio between a subject's priors in a and b be the same as the ratio between her priors in any temporal successor of a and any

Consider again the case of a subject watching a clock. Let her prior in both A(9:00) and B(9:00) be x. Since CeC assigns credences to alternatives in proportion to their priors, her credence at 9 am in A(9:00) and B(9:00) will be $\frac{1}{2}/\frac{1}{2}$. If her prior in A(9:01) and B(9:01) is also x, then her 9:01 am credences in A(9:01) and B(9:01) will also be $\frac{1}{2}/\frac{1}{2}$, as desired. But if her prior in both A(9:01) and B(9:01) was $2x$, her 9:01 am credences in A(9:01) and B(9:01) would still be $\frac{1}{2}/\frac{1}{2}$, since the ratio between their priors is the same. So to get continuity, we just need the ratio of priors between the new alternatives to be the same as the ratio of priors between the old alternatives they're continuous with.

On CoC, a subject's credences are distributed among worlds in proportion to her priors in those worlds, and her credence at each world is divided among its alternatives in proportion to her priors in those alternatives. Thus the proportion of a world's credence assigned to an alternative isn't sensitive to the absolute magnitude of the alternative's prior, only to the ratio between its prior and the priors of the other alternatives at that world. So all we need to keep track of is the ratios of the priors between alternatives at each world.[8] Thus on CoC, the Continuity Principle just requires that the ratio of priors between the new alternatives at each world be the same as the ratio of priors between any old alternatives at that world they're continuous with.

Consider again the case of a subject watching a clock, but this time let her have two alternatives at each world, A(9:00) and A$'$(9:00) at A, and B(9:00) and B$'$(9:00) at B. Let her prior in worlds A and B be y, and her prior in each of these four centered worlds be x. Since CoC assigns credences to worlds in proportion to their priors, her credence in A and B will be $\frac{1}{2}/\frac{1}{2}$ at both at 9 am and 9:01 am.

temporal successor of b.

[8] What about the ratios of priors between worlds? We don't need to put constraints on these because worlds don't get replaced by temporal successors, so the ratios between their priors are static.

Since CoC divides the credence of a world among its alternatives in proportion to their priors, her 9 am credence in each world will be split evenly between the two alternatives at that world, and her 9 am credence in each alternative will be $\frac{1}{4}$. Now, if her prior in each of the four temporal successors to these alternatives (A(9:01), A′(9:01), B(9:01) and B′(9:01)) is also x, then her 9:01 am credence in these successors will also be $\frac{1}{4}$, as desired. But if her prior in A(9:00) and A′(9:00) was $\frac{1}{2}x$, and her prior in B(9:00) and B′(9:00) was $2x$, her 9:01 am credences in these successors would still be $\frac{1}{4}$. Her credence in A and B will be $\frac{1}{2}$, and this will be divided evenly between the two alternatives at each world. So to get continuity, we just need the ratio of priors between new alternatives at each world to be the same as the ratio of priors between the old alternatives they're continuous with.

Notice that CoC requires a strictly weaker constraint on priors than CeC in order to satisfy the Continuity Principle. This is because CoC captures more of our intuitions about how our credences should be diachronically coordinated, and so requires fewer constraints on priors to keep our credences in line. In the next section we'll see why this is so, and we'll take a careful look at the extent to which CoC succeeds in capturing these intuitions.

5.3.2 Continuity and Dynamics

To what extent do CeC and CoC need a Continuity Principle in order to capture our intuitions about how our credences should evolve? The former badly needs a Continuity Principle in order to get acceptable credal behavior; without it our credences can vary arbitrarily without constraint. The latter, on the other hand, does well without a Continuity Principle. On CoC our credences in worlds aren't subject to arbitrary variation, and this limits the potential for arbitrary variation in our credences in alternatives. For subjects like us, this results in naturally coordinated credences for almost all of our alternatives. Let's look at these claims in more detail.

On CeC our credence in a doxastic world hangs on the priors of our current alternatives at that world, so as our alternatives change, our credence in the world can fluxuate wildly. On CoC our credence in a doxastic world hangs on the prior of that world, and as this value is static, our credence isn't subject to arbitrary variation. So unlike CeC, CoC naturally coordinates our credences in worlds.

Let's look at an example of how CoC coordinates our credences in worlds, and CeC does not. Consider again the subject who is watching a clock she knows to be accurate, and who has two doxastic worlds, A and B. At 9 am she has one doxastic alternative at each world, A(9:00) and B(9:00), and has equal credence in each. When she sees the clock register 9:01 am she'll replace each of her 9 am alternatives with a 9:01 am alternative.

If she's a centered conditionalizer, the fact that her 9 am credences in A(9:00) and B(9:00) were equal entails that her priors in A(9:00) and B(9:00) must be equal. But this doesn't say anything about her priors in A(9:01) or B(9:01). So if she's a centered conditionalizer her credences in the A and B worlds at 9:01 am can be completely unrelated to her credences in A and B at 9 am.

If she's a compartmentalized conditionalizer, the fact that her 9 am credences in A(9:00) and B(9:00) are equal entails that her priors in the worlds A and B are equal, although her priors in the centered worlds A(9:00) and B(9:00) may not be. This doesn't say anything interesting about her priors in A(9:01) or B(9:01), of course, but it doesn't matter.[9] Her credence in A and B at 9:01 am will be $\frac{1}{2}/\frac{1}{2}$ regardless of her priors in A(9:01) and B(9:01). So if she's a compartmentalized conditionalizer she'll naturally have coordinated credences in A and B.

This coordination of our credences in worlds substantially restricts the potential for arbitrary variation in our credences in alternatives. To see this, let's

[9] Her prior in A and B does tell us some *uninteresting* things about her priors in A(9:01) and B(9:01), of course. Since the priors of worlds are equal to the sum of the priors of the centered worlds at that world, we know that $hp(A) \geq hp(A(9{:}00)) + hp(A(9{:}01))$, for example.

look at what kinds of arbitrary variation CoC allows. At doxastic worlds with a single alternative, the alternative is assigned all of the world's credence. Since a lone alternative and its temporal successor will both be assigned the full credence of the world, and the credences of worlds are intuitively coordinated, the pair of alternatives will have intuitively coordinated credences as well. So there won't be arbitrary variation in the credence of alternatives at single alternative worlds. The only place where arbitrary variation can arise is at doxastic worlds with multiple alternatives. At multiple alternative worlds the credence of a world is divided among alternatives in proportion to their priors. If the priors of temporal successors have different relative magnitudes than their predecessors, they'll be assigned different proportions of the world's credence, and the credences of the old and new alternatives won't be intuitively coordinated.

So on CoC, arbitrary variations in the credences of alternatives can only happen at multiple alternative worlds. And unlike CeC, the amount of arbitrary variation is restricted to how the credences of worlds are divided among their alternatives.

Consider again a case where a subject is looking at a clock she knows to be accurate. As before, let her have two doxastic worlds at 9 am, A and B. This time, however, let her have one alternative at A and two alternatives at B: one centered on her, and one centered on a duplicate of her. At 9 am let her credence in A and B be $\frac{1}{2}/\frac{1}{2}$, and her credence in the two alternatives at B be $\frac{1}{4}/\frac{1}{4}$. Now, at 9:01 am the clock changes, and she replaces each of her 9 am alternatives with a 9:01 am alternative. What should her 9:01 am credences be like according to CeC and CoC?

If she's a centered conditionalizer, the ratio of her credences in her alternatives at 9 am will be the same as the ratio of her priors in those alternatives. So the ratio of her priors in A(9:00), B(9:00) and B$'$(9:00) will be 2:1:1. This doesn't entail anything about her priors in their 9:01 am successors, however, and her credence at 9:01 am in each of the 9:01 am alternatives might be anything between 0 and 1.

If she's a compartmentalized conditionalizer, having equal credence in A and B at 9 am entails that her priors in A and B are the same. Likewise, having equal credence in B(9:00) and B′(9:00) at 9 am entails that her priors in B(9:00) and B′(9:00) are the same. Her 9 am credences don't tell us anything about her prior in A(9:00), however, since her credence in A(9:00) will just be her credence in A regardless. As with CeC, none of this tells us anything interesting about her priors in her 9:01 am alternatives. But her priors in the A and B worlds will be the same, so her credence at 9:01 am in A and B will be the same as well: $\frac{1}{2}/\frac{1}{2}$. The stability of her credence in worlds imposes stability on her credences in alternatives. At single alternative worlds like A, there is no potential for arbitrary variation: the alternative at that world, A(9:01), will just be assigned the credence of that world, $\frac{1}{2}$. At multiple alternative worlds like B, there is potential for arbitrary variation. If B(9:01) and B′(9:01) have different priors, they'll be assigned different proportions of B's credence. But this isn't the extreme variation allowed by CeC; it's not the case that her credence in each alternative might be anything between 0 and 1. The only variation CoC allows is in how the credence of a world is divided among the alternatives at that world. In this case, her 9:01 am credences in B(9:01) and B′(9:01) are restricted to values between 0 and 1/2.

For subjects like us, the natural constraints on arbitrary variation imposed by CoC lead to almost complete credence coordination. There is only potential for arbitrary variation at doxastic worlds with multiple alternatives, and for subjects like us, such worlds are rare.

One way to see how rare doxastic worlds with multiple alterantives are is to note that many cases which may seem to involve multiple alternatives do not. Consider the following case. As I'm writing this, I'm wondering what time it is. When I last looked at the clock it was 6 pm, but I'm now unsure as to whether it's 7 pm or 7:05 pm. It might seem like this is a case where I now have two alternatives at each of my doxastic worlds; one located at 7 pm and another located at 7:05 pm. But there is a fact about the temporal distance between when I last looked at the clock and when I typed the sentence "As I'm writing this, I'm wondering what time it is." The doxastic alternatives where it's 7 pm are at doxastic worlds where an hour has passed between these two events, while the doxastic alternatives where it's 7:05 pm are at doxastic worlds where 65 minutes have passed between these two events. So these two alternatives aren't at the same doxastic world after all, they're at different doxastic worlds.

Here's another way to see how rare multiple alternative worlds are. Worlds at which I have multiple doxastic alternatives are worlds at which there are multiple epistemic subjects in subjective states indistinguishable from my own. Now consider my life as a sequence of time-slices. Ignore times when I've been unconscious or otherwise incapable of rational thought, and consider slices that are far enough apart to be noticeably distinct. How many of these me-slices are in subjectively indistinguishable states? If I'm in the set of worlds I think I'm probably in, none of them are. Likewise, if the world is like I think it probably is, no me-slice will be in a state indistinguishable from that of any time slice of anyone else, present, future or past.

Without the addition of a Continuity Principle, CeC does nothing to keep our credences coordinated in an intuitive manner; it allows our credences to vary arbitrarily without constraint. CoC, on the other hand, does a great deal to keep our credences coordinated. If we adopt CoC, then for the majority of our doxastic worlds—worlds at which we have a single alternative—the diachronic coordination of our credences falls right out of the dynamics. And at the rest of our doxastic worlds—strange worlds with multiple alternatives—the potential for arbitrary variation of our credences is severely constrained.

5.4 Three Accounts

We've looked at two *de se* updating rules, CeC and CoC. But those rules alone don't suffice to provide a viable account of the dynamics of *de se* beliefs. As we've just seen, we need to add some auxiliary principles to get a full account.

In what follows, we'll look at three accounts of the dynamics of *de se* beliefs that employ these rules, and examine how they apply to the sleeping beauty case. The first employs CeC, and is an attempt to yield Elga's response to the sleeping beauty case. The second also employs CeC, and is an attempt to yield Lewis's response to the sleeping beauty case. The third employs CoC, and is an account offered in Meacham (2006).[10] I'll call these accounts CeC_E, CeC_L and CoC_M, respectively. (It's worth noting that Halpern (2005) has suggested a way of capturing Elga's response that is similar in many ways to CeC_E, and Halpern and Tuttle (1993) have proposed an account of temporal belief change similar to CoC_M. I discuss these proposals, and some tentative problems they encounter, in Appendix 9.12.)

To give an intuitive feel for how these accounts work, let me briefly sketch how each treats the sleeping beauty case. Then, in the sections that follow, I'll present the accounts, and give a more detailed description of how they yield their treatment of the sleeping beauty case.

[10] With one minor difference. Meacham (2006) notes the benefits of adopting Elga's Indifference Principle, but does not go so far as to explicitly adopt it. For ease of comparison, I will take my proposal there to include Elga's Indifference Principle.

5.4.1 Sleeping Beauty

Recall the sleeping beauty case, described in section 1.4:

> *The Sleeping Beauty Case:* Some researchers are going to put Sleeping Beauty to sleep for several days. They will put her to sleep on Sunday night, and then flip a coin. If heads comes up they will wake her up on Monday morning. If tails comes up they will wake her up on Monday morning and Tuesday morning. And in between Monday and Tuesday, while she's sleeping, they will erase the memories of her waking.
>
> What should Beauty's credences be when she wakes up on Monday morning? And what should Beauty's credences become if she's told that it's Monday?

How do CeC$_E$, CeC$_L$ and CoC$_M$ deal with this case?

Elga (2000) proposes that upon waking Beauty should have a $\frac{1}{3}$ credence in heads and a $\frac{2}{3}$ credence in tails, the latter split evenly between Monday and Tuesday. If Beauty then learns that it's Monday, she should regain her original $\frac{1}{2}/\frac{1}{2}$ credence in heads/tails. This is the result CeC$_E$ yields.

Lewis (2001) proposes that Beauty retain her $\frac{1}{2}/\frac{1}{2}$ credence in heads/tails when she wakes up, with her credence in tails split evenly between Monday and Tuesday. If Beauty then learns that it's Monday, she should come to have a $\frac{2}{3}$ credence in heads and a $\frac{1}{3}$ credence in tails. This is the result CeC$_L$ yields.

Like Lewis, Meacham (2006) proposes that Beauty retain her $\frac{1}{2}/\frac{1}{2}$ credence in heads/tails when she wakes up, with her credence in tails split evenly between Monday and Tuesday. This account diverges from Lewis's with respect to what happens when Beauty learns that it's Monday. CoC$_M$ entails that her credences in heads/tails should remain $\frac{1}{2}/\frac{1}{2}$.

More generally, consider a subject with multiple doxastic worlds who undergoes a belief change that just increases or decreases (to a minimum of 1) the number of alternatives at some world. We can capture the flavor of these three accounts by looking at how such a belief change affects the subject's credence in that world. On CoC$_M$ the subject's credence will remain unchanged. On CeC$_L$, if the number of alternatives at that world increases then the subject's credence will remain unchanged. But if the number of alternatives at that world decreases, then the subject's credence will decrease as well. On CeC_E, the subject's credence will change in both cases. If the number of alternatives at that world increases or decreases, then the subject's credence in that world will likewise increase or decrease.

5.5 CeC_E: Elga's Response

In Elga's (2000) discussion of the sleeping beauty case, he proposes that after waking up Beauty's credence in heads/tails should be $\frac{1}{3}/\frac{2}{3}$, the latter split evenly between Monday and Tuesday. If Beauty then learns that it's Monday, he proposes that her credence in heads/tails should become $\frac{1}{2}/\frac{1}{2}$. Elga's response follows from three principles:

1. Centered Conditionalization
2. Elga's Indifference Principle
3. A Continuity Principle

I'll take CeC_E to be the conjunction of these three principles.

In standard treatments of the case, a chance-credence principle is also employed. But it only plays the superficial role in the argument of setting our credence in heads and tails on Sunday to $\frac{1}{2}/\frac{1}{2}$. The interesting features of the case remain regardless of Beauty's reason for having $\frac{1}{2}/\frac{1}{2}$ credences in heads and tails on Sunday. Since, as we saw in chapter 2, it's easy to get confused about the role of the chance-credence principle in the sleeping beauty case, I'll take Beauty's initial $\frac{1}{2}/\frac{1}{2}$ credence in heads/tails to be a fixed part of the case, and leave the chance-credence principle out of it.

We're familiar with the first of the three principles CeC$_E$ consists of, CeC. What about the other two?

The second principle is Elga's Indifference Principle. Elga (2004) characterizes the principle as the requirement that subjectively indistinguishable alternatives at the same world be assigned the same credence. Since all of one's alternatives are subjectively indistinguishable, this entails that all of a subject's alternatives at the same world should have the same credences. Weatherson (2005) has offered several criticisms of Elga's Indifference Principle. I discuss these criticisms, and the general plausibility of the principle, in Appendix 9.13.

For reasons I describe in Appendix 9.13, I'll formulate the principle in a slightly different way than Elga. I'll take Elga's Indifference Principle to be the following constraint:[11]

$$cr_e(\cdot|W_i) = m_i(\cdot|e), \text{ if } cr_e(\cdot|W_i) \text{ is defined,} \tag{5.5}$$

where m_i is the canonical self-locating measure over W_i posited by EF3 in section 4.4. A consequence of (5.5) is that if a subject has only finitely many alterantives at a world, her credence in each of them should be the same.

The third principle is a Continuity Principle. As we've seen, the content of a Continuity Principle depends on when we take pairs of alternatives to be continuous. To get Elga's response, any Continuity Principle for which the following is a sufficient condition for continuity will do: if a is a temporal successor of b—$S_t(a,b)$—then a and b are continuous.[12]

[11] For CeC and CoC, this is equivalent to the following constraint on priors:

$$hp(\cdot|W_i) = m_i(\cdot), \text{ if } hp(\cdot|W_i) \text{ is defined.} \tag{5.4}$$

[12] If we take this condition and replace the temporal successor relation S_t with the temporal continuant relation C_t, we can derive Elga's response to the original sleeping beauty case (but not to duplication versions of the sleeping beauty case) without employing Elga's Indifference Principle.

Elga's proposal follows from these three principles. Let $\text{cr}(\cdot)$ be Beauty's credence function and $\text{hp}(\cdot)$ her hypothetical priors. Let H/T be the propositions that the coin comes up heads/tails, and SUN/MON/TUE be the centered propositions that it's Sunday, Monday and Tuesday, respectively.

Beauty's credences in her heads and tails alternatives on Sunday will be $\text{cr}(\text{H}\wedge\text{SUN}) = \text{cr}(\text{T}\wedge\text{SUN}) = \frac{1}{2}$. Given CeC, this entails that $\text{hp}(\text{H}\wedge\text{SUN}) = \text{hp}(\text{T}\wedge\text{SUN})$. When she wakes up on Monday, her Sunday alternatives are replaced by Monday alternatives at the heads worlds, and by Monday and Tuesday alternatives at the tails worlds. At both the heads and the tails worlds, her Monday morning alternatives will be the temporal successors of her Sunday night alternatives. So according to the Continuity Principle given above, her Monday alternatives are continuous with her Sunday alternatives. We saw in section 5.3 that given CeC, the Continuity Principle requires that the ratios of priors between the new and old continuous alternatives be the same. Since $\text{hp}(\text{H}\wedge\text{SUN}) = \text{hp}(\text{T}\wedge\text{SUN})$, it follows that $\text{hp}(\text{H}\wedge\text{MON}) = \text{hp}(\text{T}\wedge\text{MON})$. Elga's Indifference Principle requires that her credences in the two alternatives at the tails worlds be equal, and given CeC this entails that $\text{hp}(\text{T}\wedge\text{MON}) = \text{hp}(\text{T}\wedge\text{TUE})$. Putting these equalities together, we get $\text{hp}(\text{H}\wedge\text{MON}) = \text{hp}(\text{T}\wedge\text{MON}) = \text{hp}(\text{T}\wedge\text{TUE})$. When she wakes up her doxastic alternatives are H$\wedge$MON, T$\wedge$MON and T$\wedge$TUE, so on CeC her credences after waking on Monday are $\text{cr}(\text{H}\wedge\text{MON}) = \text{cr}(\text{T}\wedge\text{MON}) = \text{cr}(\text{T}\wedge\text{TUE}) = \frac{1}{3}$.

Now, say she wakes up at 9 am. What if at 9:01 am she learns that it's Monday? After learning it's Monday she will have one alternative at each world, H$\wedge$MON(9:01) at the heads worlds and T$\wedge$MON(9:01) at the tails worlds. If H$\wedge$MON(9:01) and T$\wedge$MON(9:01) are the temporal successors of H$\wedge$MON(9:00) and T$\wedge$MON(9:00), then the Continuity Principle entails that they're continuous with their temporal predecessors. Since $\text{hp}(\text{H}\wedge\text{MON}(9{:}00)) = \text{hp}(\text{T}\wedge\text{MON}(9{:}00))$, it follows that $\mathrm{hp}(\mathrm{H}\wedge\mathrm{MON}(9{:}01)) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}(9{:}01))$. So on CeC her credence after learning it's Monday is evenly split between heads and tails: $\mathrm{cr}(\mathrm{H}\wedge\mathrm{MON}(9{:}01)) = \mathrm{cr}(\mathrm{T}\wedge\mathrm{MON}(9{:}01)) = \frac{1}{2}$. (If H∧MON(9:01) and T∧MON(9:01) aren't the temporal successors of H∧MON(9:00) and T∧MON(9:00), there will be pairs of alternatives at times between the two which lead, through a chain of continuity, to the same result.[13])

So these three principles yield Elga's response to the sleeping beauty case: Beauty's credence in heads/tails when she wakes up should be $\frac{1}{3}/\frac{2}{3}$ (the latter evenly split between Monday and Tuesday), and her credence in heads/tails when she is told it's Monday should be $\frac{1}{2}/\frac{1}{2}$.

5.6 CeC$_L$: Lewis's Response

In Lewis's (2001) discussion of the sleeping beauty case, he proposes that after waking up Beauty's credence in heads/tails should be $\frac{1}{2}/\frac{1}{2}$, the latter split evenly between Monday and Tuesday. If Beauty then learns that it's Monday, he proposes that her credence in heads/tails should become $\frac{2}{3}/\frac{1}{3}$. Lewis's response follows from four principles:

1. Centered Conditionalization
2. Elga's Indifference Principle
3. A Continuity Principle

[13] Isn't there also temporal succession between her T∧MON alternative before she is put to sleep and her T∧TUE alternative after she wakes up? And won't the Continuity Principle then place some additional constraint on her priors? Yes, there will be a Monday alternative that's the temporal predecessor of her first T∧TUE alternative, and yes, the Continuity Principle will apply. But in this case, its effects won't be very noticeable. If she has only one doxastic tails world, then the Continuity Principle will place no constraint on her priors. If she has multiple doxastic tails worlds, then the Continuity Principle will require that the ratios between her priors in these worlds after she wakes up is the same as the ratios between her priors in these worlds before she went to sleep. And since we're not concerned here with how her credence in T∧TUE is divided among her doxastic tails worlds, this has no bearing on our treatment of the case.

4. The No-Increase Principle

I'll take CeC$_L$ to be the conjunction of these four principles.

The first two premises are familiar. The third premise is another Continuity Principle. Although Lewis must reject CeC$_E$'s Continuity Principle, a similar principle will work. Lewis needs a Continuity Principle for which the following is a sufficient condition for continuity: if *a* is a temporal successor of *b*, and the number of alternatives at that world has not increased, then *a* and *b* are continuous.

This leaves us with the question of what constraints, if any, should be imposed on a subject's credences at worlds where the number of alternatives increases. To get Lewis's result, we want it to be the case that in cases where a subject doesn't suffer from memory loss and doesn't get evidence about the world—where she doesn't gain or lose doxastic worlds—increases in the number of alternatives at a world leave her credence in that world unchanged.[14] I'll call this the No-Increase Principle.

Lewis's response follows from these four principles, As before, we assume Beauty's credences in heads and tails alternatives on Sunday will be cr(H∧SUN) = cr(T∧SUN) = $\frac{1}{2}$. CeC then entails that hp(H∧SUN) = hp(T∧SUN). When she wakes up on Monday, her Sunday alternatives are replaced by Monday alternatives at the heads worlds, and by Monday and Tuesday alternatives at the tails worlds. By the No-Increase Principle the increase in alternatives at her tails worlds should leave her credence in tails unchanged, so her credence in tails after waking up on Monday is the same as her credence in tails on Sunday, $\frac{1}{2}$. So her credence

[14] In Meacham (2006) I left out the "and doesn't suffer from memory loss" clause. This clause is required to keep CeC$_L$ from being inconsistent. If Beauty learns it's Monday on Monday, her credence in heads/tails on CeC$_L$ will become $\frac{2}{3}/\frac{1}{3}$. And if the coin lands tails, then CeC requires her credence in heads/tails on Tuesday morning to be the same as on Monday morning—$\frac{1}{2}/\frac{1}{2}$—since she will have the same evidence. But without the memory loss clause, the No Increase Principle would apply and require her credence in heads/tails to also be the same on Tuesday morning as it was on Monday night—$\frac{2}{3}/\frac{1}{3}$—which leads to a contradiction.

in heads after waking up on Monday must be $\frac{1}{2}$ as well. Given CeC, this entails that hp(H∧MON) = hp(T∧(MON∨TUE)) = hp(T∧MON) + hp(T∧TUE). Elga's Indifference Principle and CeC entail that hp(T∧MON) = hp(T∧TUE). Taken together, these equalities entail hp(H∧MON) = hp(T∧MON) + hp(T∧TUE) = 2·hp(T∧MON) = 2·hp(T∧TUE). When she wakes up her doxastic possibilities are H∧MON, T∧MON and T∧TUE, so on CeC her credences after waking up on Monday are cr(H∧MON) = $\frac{1}{2}$ and cr(T∧MON) = cr(T∧TUE) = $\frac{1}{4}$.

Now what if Beauty is woken up at 9 am and told at 9:01 am that it's Monday? After learning it's Monday she will have one alternative at each world. If these 9:01 am alternatives are temporal successors of her Monday 9 am alternatives, then they will be continuous. We know from above that hp(H∧MON(9:00)) = 2·hp(T∧MON(9:00)), so it follows that hp(H∧MON(9:01)) = 2·hp(T∧MON(9:01)). So on CeC her credences after learning it's Monday are cr(H∧MON(9:01)) = $\frac{2}{3}$ and cr(T∧MON(9:01)) = $\frac{1}{3}$. (If her 9:01 am alternatives are not temporal successors of her 9 am alternatives, there will be pairs of alternatives at times between the two which are continuous, and which will lead to the same result.)

So these four principles yield Lewis's response to the sleeping beauty case: Beauty's credence in heads/tails when she wakes up should be $\frac{1}{2}/\frac{1}{2}$ (the latter evenly split between Monday and Tuesday), and her credence in heads/tails when she's told it's Monday should be $\frac{2}{3}/\frac{1}{3}$.

5.7 CoC$_M$: A Third Response

In Meacham's (2006) discussion of the sleeping beauty case, I propose that after waking up Beauty's credence in heads/tails should be $\frac{1}{2}/\frac{1}{2}$, the latter split evenly between Monday and Tuesday. If Beauty then learns that it's Monday, he proposes that her credence in heads/tails should remain $\frac{1}{2}/\frac{1}{2}$. This response follows from two principles:

1. Compartmentalized Conditionalization
2. Elga's Indifference Principle

I'll take CoC_M to be the conjunction of these two principles.

It might be somewhat surprising that we haven't included a Continuity Principle on this list. In section 5.3 we learned that both CoC and CeC require some kind of prior constraint in order for our credences to be diachronically coordinated in the appropriate way. But a much weaker constraint on priors is required if we adopt CoC rather than CeC. We can see how weak this constraint is by noting that, given CoC, Elga's Indifference Principle imposes a strictly stronger constraint on priors than the Continuity Principle. Recall that given CoC, the Continuity Principle requires that the ratio of priors between new alternatives at each world be the same as the ratio of priors between any old alternatives at that world that they're continuous with. If you adopt Elga's Indifference Principle, then your credences in alternatives at a world will be the same, and thus so will your priors. If your priors in alternatives at a world are always the same, the ratio of priors between alternatives at a world will always be 1:1, and the Continuity Principle will be automatically satisfied. So on CoC, a person who adopts Elga's Indifference Principle needn't adopt any further principles in order to get completely coordinated credences.

Meacham's (2006) response follows from these two principles. On CoC a subject first divides her credences among worlds, and then divides the credence of each world equally among the alternatives at that world. So a subject's credence in worlds, and thus in *de dicto* propositions, only changes when she gains or loses doxastic worlds.

On Sunday Beauty has a $\frac{1}{2}/\frac{1}{2}$ credence that the coin toss came up heads/tails, with one doxastic alternative at each of her doxastic worlds. When she wakes up on Monday she has one alternative (Monday) at each heads world and two alternatives (Monday and Tuesday) at each tails world. But although her doxastic alternatives have changed, she has the same doxastic worlds she had on Sunday. Since her doxastic worlds have remained the same, she will have the same credence in heads/tails: $\frac{1}{2}/\frac{1}{2}$. How should her $\frac{1}{2}$ credence in tails be divided between Monday and Tuesday? Elga's Indifference Principle requires this to be split evenly between Monday and Tuesday at the tails world. So Beauty's credences after waking up on Monday will be cr(H∧MON) $= \frac{1}{2}$ and cr(T∧MON) = cr(T∧TUE) $= \frac{1}{4}$.

What if she then learns that it's Monday? This eliminates the Tuesday alternative at her tails worlds, but doesn't eliminate any doxastic worlds. So again, her credence in heads/tails will remain the same: $\frac{1}{2}/\frac{1}{2}$.

So these two principles yield Meacham's (2006) response to the sleeping beauty case: Beauty's credence in heads/tails when she wakes up should be $\frac{1}{2}/\frac{1}{2}$ (the latter evenly split between Monday and Tuesday), and her credence in heads/tails when she's told it's Monday should remain $\frac{1}{2}/\frac{1}{2}$.

5.8 Worries for CeC$_E$

Now let us turn to assessing the three accounts. In the next three sections, I'll point out some worries one might have about each of these accounts, and in the final section, I'll evaluate their prospects.

Let's start with the account that yields Elga's response, CeC$_E$.

5.8.1 The Information About the World Worry

There's a clear sense in which Beauty doesn't learn anything new about the world between the time she goes to sleep on Sunday and the time she wakes up on Monday: she doesn't get information which eliminates doxastic worlds. Elga's response to the sleeping beauty case has struck many as strange because it recommends changes in Beauty's credences about the world, even though it seems she hasn't gained or lost any information about what the world is like. And CeC$_E$ inherits this worry. On CeC$_E$, when Beauty wakes up on Monday her credence in whether the coin landed heads changes—it becomes $\frac{1}{3}$ instead of $\frac{1}{2}$—even though she has the same doxastic worlds as she had on Sunday.

It's contentious how much weight we should give this objection. Most will agree that the following maxim is plausible: "your credences about the world shouldn't change if you don't get any information about what the world is like." But CeC$_E$'s recommendations only violate this maxim if we understand "don't get any information about what the world is like" to mean "don't get evidence which eliminates doxastic worlds". Suppose we take "don't get any information about what the world is like" to mean "don't get evidence relevant to one's credence in *de dicto* propositions". Then CeC$_E$ will claim that Beauty *is* getting information about the world when she wakes up on Monday. After all, when Beauty wakes up on Monday her credence that the coin landed heads should change from $\frac{1}{2}$ to $\frac{1}{3}$. So according to the characterization just given, she *has* gotten information about what the world is like. And to simply deny this is to beg the question against CeC$_E$.

5.8.2 The Continuity Worry

Another worry about CeC$_E$ is that the updating rule it employs—CeC—requires a Continuity Principle in order to provide a satisfactory account of how our credences at different times ought to be related. If CeC were the correct updating rule, one might argue, it should entail that a subject's credences are diachronically coordinated in the appropriate way by itself. After all, that's what an updating rule is supposed to be giving us.

Note that this complaint doesn't apply to other prior-constraining principles, like the chance-credence relation. An updating rule isn't flawed if it doesn't encode the chance-credence relation. An updating rule is just supposed to capture how our credences at different times should be related, and it doesn't seem like the chance-credence relation has much to do with that. But if our rule requires a Continuity Principle to get an acceptable account of how a subject's credences should be diachronically coordinated, then it seems we might have the wrong updating rule.

(Can't we just call the conjunction of CeC and a Continuity Principle the "updating rule", and sidestep this worry? No: an updating rule is just a diachronic credence constraint. An updating rule by itself shouldn't constrain which kinds of initial credence functions are rationally permissible. Since the conjunction of CeC and a Continuity Principle imposes such constraints, it can't plausibly be called an updating rule. (Halpern (2005) has proposed some rules that fall into this category. In Appendix 9.12, I discuss these rules, how they avoid Continuity Principles, and the potential problems they run into as a result.))

I take this to be a compelling worry for CeC. But this is a worry that comes in degrees. An updating rule which captures most of our intuitions about how credences should be diachronically coordinated is better than a rule which captures fewer of these intuitions. And perhaps CeC will do well enough to remain a plausible candidate for the correct updating rule.

5.8.3 The Many Brains Argument

The following case brings out a third worry for CeC$_E$:

> *The Many Brains Argument:* Consider the hypothesis that you're a brain in a vat. Although this is epistemically possible and (perhaps) nomologically possible, your current credence in this possibility is presumably very low. Now consider the proposition that you're in a world where brains in vats are constantly being constructed in states subjectively indistinguishable from your own. Let your credence in this proposition be $0 < p < 1$, and your credence that there will be no multiplication of doxastic alternatives be $1-p$. If you accept CeC$_E$ then your credence in this hypothesis will increase over time and converge to 1. Thus you should come to believe (if not yet, then

> in a little while) that these brains in vats are being created. (A proof of this result is provided in appendix 9.14.)

It follows from Elga's Indifference Principle that your credences should be spread evenly among the doxastic alternatives at a world. So as you become certain that these brains in vats are being created, you should become certain that you're a brain in a vat.

The many brains argument assumed that brain in a vat duplication is the only proposition in which you have a non-zero credence that multiplies doxastic alternatives. But the result generalizes. Suppose that you also have a small credence in the proposition that you're in a world where duplicates of you are constantly being created on distant but qualitatively identical planets. Then you'll come to believe (if not yet, then in a little while) that these brains in the vats are being created *or* that these duplicates of you are being created. Likewise, you'll come to believe that you are a brain in a vat *or* a duplicate on a distant planet. By a similar process, you can generalize the result of the many brains argument to any number of propositions that multiply alternatives.

In general, if you accept CeC$_E$ then you will come to believe that you're in a world where you have many doxastic alternatives. These are strange worlds. So if we accept CeC$_E$, we'll come to believe (if not yet, then in a little while) that we live in a strange world. This is an unwelcome consequence.

5.9 Worries for CeC$_L$

Now let's turn to the account that yields Lewis's response, CeC$_L$.

The first two worries about CeC$_E$ given above apply to CeC$_L$ as well. Like CeC$_E$, CeC$_L$ recommends changes in Beauty's credences about the world, even though she hasn't gained or lost any information about what the world is like. CeC$_L$, when Beauty learns that it's Monday, her credence in heads changes—it becomes $\frac{2}{3}$ instead of $\frac{1}{2}$—even though she has the same doxastic worlds. Likewise, CeC$_L$ employs CeC as well, and so requires a Continuity Principle to provide a satisfactory account of how our credences at different times ought to be related.

In addition, CeC$_L$ faces the following worries.

5.9.1 The Asymmetry Worry

On CeC$_L$, increasing the number of alternatives at a world will generally have no effect on your credence in that world, but decreasing the number of alternatives will generally decrease your credence in that world. This seems odd. What justifies this asymmetric treatment of alternative multiplication and elimination? Without some some further justification for this behavior, the asymmetry of CeC$_L$ seems arbitrary. Since the two principles that encode this asymmetry—CeC$_L$'s Continuity Principle and the No-Increase Principle—don't seem particularly natural or well-motivated, CeC$_L$ looks hopelessly *ad hoc*.

This criticism isn't necessarily fatal. If an updating rule does a good enough job of accounting for our intuitions, the fact that it looks *ad hoc* can be overcome. Alternatively, a proponent of CeC$_L$ might be able to come up with a plausible account for why our credences in alternatives should be asymmetric in this way. But as things stand, I take this to be a reason to dislike CeC$_L$.

5.9.2 The Sadistic Scientists Argument

CeC$_E$ was subject to the many brains argument because it entailed that belief changes that multiply alternatives at a world generally increase one's credence in that world. CeC$_L$ avoids this result by adopting a different Continuity Principle and the No-Increase Principle. But while on CeC$_L$ account belief changes that multiply alternatives at a world don't increase one's credence in that world, belief changes that decrease the number of alternatives at a world generally do decrease one's credence in that world. Consider the following case:

> *The Sadistic Scientists Argument:* Consider the hypothesis that you're in a world where every second some scientists will create n brains in vats in situations subjectively identical to your own. A half second after the brains are created, the scientists will destroy them. Let your credence in this proposition be $0 < p < 1$, and your credence that there will be no creation or destruction of doxastic alternatives be $1 - p$. When the brains are created your credence that you are in such a world will remain the same (No-Increase Principle), and this credence will be evenly split between your $n + 1$ alternatives (Indifference Principle). As a half second passes and these brains are destroyed, your credence that you are in such a world will decrease by the appropriate amount (Continuity Principle and centered conditionalization). So as each second passes, your credence that you are in such a world will decrease and converge to 0. Thus, if you hold CeC$_L$ you should come to believe (if not yet, then in a little while) that these brains in vats are not being created. (A proof of this result is provided in appendix 9.15.)

The sadistic scientists argument assumed that brain in vat destruction is the only proposition you have a non-zero credence in that diminishes alternatives. Now suppose that you also had a small credence in the proposition that duplicates of you on distant but qualitatively identical worlds were being created and destroyed. Then you'd come to believe (if not yet, then in a little while) that neither of these propositions was true. The result generalizes to any number of propositions that diminish alternatives. In general, if you accept CeC$_L$ then you'll come to believe that you're not in a world where continual doxastic elimination is taking place.

I take this result to be counterintuitive. If the result as stated does not move you, imagine a case in which you are living in a world where brain-in-a-vat creation technology is cheap and easily accessible. An enemy of yours who would enjoy destroying brains in vats in your subjective state tells you that at midnight she'll spend an hour creating n such brains, and at 1 am she'll spend an hour destroying them. This enemy has the resources to carry out this threat, and reliably carries out the threats she makes. If n is big enough, and you hold CeC$_L$, then though you're now almost certain that she will carry out her threat, when you wake up tomorrow morning you'll be almost certain that she didn't. Indeed, if n is big enough, you could even go with her and watch as she creates the brains and destroys them; if you watch for long enough you won't believe your eyes!

5.10 Worries for CoC$_M$

Finally, let's turn to the third account, CoC$_M$. CoC$_M$ is also subject to a version of the continuity worry I raised for CeC$_E$ and CeC$_L$. CoC$_M$ employs CoC as an updating rule instead of CeC, but CoC also requires a Continuity Principle in order to provide a satisfactory account of how our credences at different times ought to be related. That said, CoC seems to be better of in this respect than CeC. Without a Continuity Principle, CoC still provides the desired diachronic coordination of our credences at normal worlds where we have a single alternative. And at doxastic worlds with multiple alternatives, CoC imposes a strong restriction on the potential for arbitrary variation. Likewise, CoC requires a much weaker Continuity Principle than CeC does. Given CoC, we can get all of the continuity we need if we adopt what is arguably an independently plausible principle, Elga's Indifference Principle.

But CoC$_M$ faces other worries as well.

5.10.1 The Worlds/Centered Worlds Divide

On CeC, worlds are not special. Worlds are treated in exactly the same way as any other set of centered worlds. On CoC, on the other hand, worlds are special. Beliefs about what the world is like are treated differently from self-locating beliefs. While in some cases our intuitions seem to respect this divide, in others cases the distinction is less clear. By homing in on cases where this distinction is not salient, we can construct a class of cases where CoC appears to deliver counterintuitive results.

There are two kinds of cases one can construct this way.

(A) The first leans on the way in which CoC doesn't 'count' multiple centered worlds when it evaluates what your credences in worlds should be. In some cases, this is clearly a boon: this is how CoC avoids the Many-Brains arguments that afflict CeC. But one can also try to set up cases where our intuitions swing the other way.[15]

(B) The second constructs a pair of cases—one self-locating and one not—which are intuitively similar. In general, CoC will treat the two cases differently, and one can argue that this is counterintuitive.[16]

Let's look at examples of each of these two kinds of cases.

(A) Why We Hate Mondays

> *Why We Hate Mondays:* Consider a case like the sleeping beauty case, but where the awakening process is slightly different. In this case, when they wake you up on Monday, they'll do so by blaring loud music that you hate. If they wake you up on Tuesday, they'll do so by softly playing music that you love. Now say you wake up to loud music that you hate. What should your credence be in heads/tails?

Hearing loud music you hate doesn't eliminate any of the heads or tails worlds, so according to CoC, your credences should remain $\frac{1}{2}/\frac{1}{2}$. According to CeC, on the other hand, your credence in heads should increase to $\frac{2}{3}$.

One might think the response CoC gives is counterintuitive. To amplify the intuition, consider a variant of the above case where, if tails comes up, you'll be woken up once a day for a year, with soft music you love being used on every day but the first. When you wake up to loud music, isn't it unlikely that tails came up? After all, if tails came up there would be 364 soft music awakenings, and you'd have to be pretty unlucky to not have gotten one of them.

[15] People who have argued that this can also yield counterintuitive consequences include (in chronological order) David Manley, Matt Kotzen and Sarah Moss.

[16] This sort of case has been raised by a number of people, including (in chronological order) David Manley, Matt Kotzen, Robert Stalnaker and Sarah Moss.

While it's easy to get the feel of this objection, it's hard to cash it out in a coherent manner. I find that when I try to spell out the intuition behind this kind of reasoning, it stops looking so appealing. Consider again the last two sentences of the last paragraph: "When you wake up to loud music, isn't it unlikely that tails came up? After all, if tails came up there would be 364 soft music awakenings, and you'd have to be pretty unlucky to not have gotten one of them." In terms of objective probability, you're as "likely" to hear loud music given tails as heads. And since you're guaranteed to be woken up to loud music no matter what, it's hard to make sense of the way in which you'd be "unlucky" to wake up to loud, blaring music if the coin came up tails. And it's not clear what other notions of "likely" and "unlucky" one could employ in order to make this reasoning coherent. Perhaps there are other ways of cashing out the intuitions in question, but in the absence of such, I find it hard to give this kind of example much weight.

(B) Sleeping Beauty and Sleeping Cutie

Consider two cases. First, the standard sleeping beauty case:

> *The Sleeping Beauty Case:* Some researchers are going to put Beauty to sleep for several days. They will put her to sleep on Sunday night, and then flip a coin. If heads comes up they will wake her up on Monday morning. If tails comes up they will wake her up on Monday morning and Tuesday morning, and in-between Monday and Tuesday, while she is sleeping, they will erase the memories of her waking. (Regardless of the outcome of the coin toss, or the day of awakening, things will be set up so that each awakening is subjectively identical.)

Second, the following variant of the sleeping beauty case:

> *The Sleeping Cutie Case:* Some researchers are going to put Cutie to sleep for several days. They will put her to sleep on Sunday night, and then flip a coin. If heads comes up they will wake her up on Monday morning. If tails comes up they will wake her up on Monday morning and Tuesday morning, and in-between Monday and Tuesday, while she is sleeping, they will erase the memories of her waking.
>
> However, in this case, the awakenings needn't be subjectively identical. In particular, let's say, the window will be open when she wakes up, and she

> will be able to see what the weather is like outside. (Further suppose that she can distinguish between 100 different kinds of weather, and that she knows each is equally probable.)

On CoC a subject first divides her credences among worlds, and then divides the credence of each world equally among the alternatives at that world. In the first case, the doxastic worlds of the subject are the same on Sunday and Monday, so CoC will recommend that her credence in heads and tails upon waking up should remain $\frac{1}{2}/\frac{1}{2}$. In the second case, however, the doxastic worlds of the subject do change. Before she went to sleep she didn't know what the weather would be like on Monday (or Tuesday). When she wakes up, however, she sees a particular kind of weather outside, and will eliminate those worlds incompatible with this evidence. More precisely, she will eliminate the $\frac{99}{100}$ of the heads worlds where the Monday weather doesn't match what she sees, and she'll eliminate the $\frac{9801}{10000}$ of the tails worlds where neither the Monday nor the Tuesday weather matches what she sees. (These fractions are the fractions of her overall credence in the heads and tails worlds, respectively.) As a result CoC will recommend that her credence in heads and tails upon waking up be $\frac{100}{10000}/\frac{199}{10000}$, or approximately $\frac{1}{3}/\frac{2}{3}$. (As we increase the number of distinguishable kinds of weather, this value converges to $\frac{1}{3}/\frac{2}{3}$.)

But, one might argue, aren't the sleeping beauty and sleeping cutie cases similar in relevant respects? Shouldn't our credences in each case be at least approximately the same?

I think there are two distinct worries behind this objection. To disentangle them, let us first look at how a proponent of CoC might respond.

> The sleeping beauty case and the sleeping cutie case are not similar in relevant respects. Beauty is not learning anything about the world—when she wakes up she has the same doxastic worlds as before—so her credences in *de dicto* propositions should remain the same. But when Cutie sees what the weather is like she is learning something about the world—when she wakes up she eliminates a number of her old doxastic worlds—so her

> credences in *de dicto* propositions should change. So the results that CoC provides are what we should expect.

One might be unhappy with this response because while it's true that Cutie learns something about the world, it doesn't seem she's learning anything about the world that should result in her changing her credences in heads/tails. Here's the first way to press this worry:

> Although Cutie learns what the weather is like, this doesn't seem to have anything to do with the outcome of the coin toss. CoC requires Cutie's credence in heads/tails to change in the same way regardless of what kind of weather she sees. But then it's hard to see how seeing a particular kind of weather could justify this change in her beliefs.

This is a deep and interesting worry. We'll look at a case which isolates and magnifies this worry—the Varied Brains Arguments—below.

Here's a second way to press this worry:

> In the sleeping cutie case we have a pair of probabilistically independent chance processes: the coin toss of the scientists, and the process that determines what the weather will be like. And once you see what the weather is like, CoC recommends that you change your credences in heads/tails. But isn't it crazy to think that evidence about a process probabilistically independent of the coin toss should have a bearing on your credence in the outcome of the coin toss? How could evidence about the one serve as evidence for the other, when the two are independent? Since CoC has this consequence, CoC must be wrong.[17]

Let's get clear about what seems crazy here. Consider two independent probabilistic processes, X and Y. Let the outcomes of the first process be $X_1, ..., X_n$, and the outcomes of the second process be $Y_1, ..., Y_m$.[18] Given that one's credence function respects this independence, it does indeed seem crazy to think that learning X_1 should have a bearing on your credence in the Ys. After all, $cr(Y_j|X_1) = cr(Y_j)$, so if you update by something like conditionalization then your credences should remain the same.

[17] This worry was raised by David Manley.

[18] So $\forall i, j\ (p(X_i \wedge Y_j) = p(X_i)p(Y_j))$.

But it doesn't follow from this that any evidence about the outcome of X should have no bearing on your credence in Ys. After all, while the Xs themselves are probabilistically independent of the Ys, the means by which you get your evidence about the Xs may not be. We can only be confident that our evidence about X should have no bearing on our credence in Y if our evidence is about the outcomes of X alone. (I.e., if the evidence can be expressed as a Boolean construction out of propositions expressing X outcomes alone.) If this isn't the case, then we shouldn't expect your credence in Y outcomes to remain unchanged.

Say, for example, you employ a stock broker to take care of your assets. Given past experience, you know that your earnings are as likely to go up as down in any given month, and that the stock broker is as likely to be on vacation as not in any given month. Furthermore, you know that the likelihood of your earnings going up or down is (sadly) independent of the likelihood of your stock broker being on vacation. Finally, you know that your stock broker only calls to tell you how things are going when your earnings have just gone up and he's not on vacation. Now, suppose he calls you to tell you that your earnings have gone up. This is evidence about your earnings, and your beliefs about your earnings will change in response to it. But this evidence will also have a bearing on your beliefs about whether he's on vacation, even though your earnings and his vacationing are probabilistically independent. This is because what evidence you get about your earnings is not independent of his vacationing, even though your earnings themselves are.

Now, what about the sleeping cutie case? Although the weather is probabilistically independent of the coin toss, what evidence we get about the weather is not independent of the outcome of the coin toss. Consider the possible weather patterns on Monday and Tuesday, divided as finely as you can distinguish. This gives us 10,000 possibilities, each equally likely. Now, say we wake up and see that it's sunny. Then we'll have learned that either it's sunny on Monday (if heads came up) or that it's sunny on Monday or Tuesday (if tails came up). But there's no way to express this proposition if we restrict ourselves to the Boolean combinations of propositions only expressing possible weather outcomes. We can't just say that our evidence is "it's sunny on Monday or Tuesday", for example, because we've learned more than that: we've also learned that if the coin toss came up heads, it's sunny on Monday.

So the probabilistic independence objection is misguided. When you see the weather, you aren't, in fact, learning something about a coin toss-independent process, in a coin toss-independent way. You're learning something about a coin toss-independent process in a coin toss-*dependent* way. And it's not counterintuitive that this could change your credence in the outcome of the coin toss!

5.10.2 The Varied Brains Argument

> *The Varied Brains Argument:* Assume that your doxastic worlds are such that they can be divided into two kinds of worlds, normal worlds and strange worlds. Throughout your doxastic worlds there are n subjectively distinguishable experiences that you might experience in the next second. Assume that you have some normal doxastic world compatible with each experience, and you have no subjective duplicates at your normal doxastic worlds. Assume that at each of your strange doxastic worlds there are scientists who will create n brains in vats a second from now, each brain compatible with one of your possible experiences. Now, at the end of a second you'll have some experience, say that of eating chocolate ice cream. This will eliminate the many normal worlds in which you don't have the experience of eating chocolate ice cream. On the other hand, at all of your strange worlds there's a brain in a vat which has the experience of eating chocolate ice cream, so no strange worlds will be eliminated. On CoC, your credence in your strange doxastic worlds should increase relative to your credence in your normal doxastic worlds.

We can extend this case by replacing 'second' with longer units of time, and as the unit of time grows larger, the number n of distinguishable experiences you might experience grows larger as well. By making the unit of time arbitrarily large, we can get a case in which, on CoC, your credence in your strange doxastic worlds grows arbitrarily large. This is an unwelcome consequence.

In Meacham (2006) I argued that this result was not as bad as the Many Brains arguments because it's less likely to afflict agents like us. But it's not clear that even this is true. As Tim Maudlin has pointed out, we are confronted with a version of the Varied Brains argument when we evaluate the Many Worlds interpretation of quantum mechanics versus (say) Bohm's theory. Suppose, for simplicity, that we know the initial wave function of the universe, and that our credence is split evenly between Many Worlds and Bohm's theory. Whenever we get evidence that depends on the outcome of a probabilistic quantum mechanical process, we will eliminate a number of Bohmian doxastic worlds—the worlds where the probabilistic process resulted in a different outcome. At the same time, we will eliminate a number of Many Worlds branches—centered worlds—from our Many Worlds doxastic world. On CoC, this will result in our credence in Bohm's theory decreasing and our credence in Many Worlds increasing, since some Bohmian worlds were eliminated, and no Many Worlds worlds were. And as we get more and more evidence, our credence in Many Worlds will continue to increase until we're virtually certain that Many Worlds is the correct interpretation of quantum mechanics.

So the varied brains result seems as bad as the many brains result. Both apply to agents like us, and both have highly counterintuitive consequences.

5.11 Assessing the Three Accounts

So what should we think about these three accounts?

In Meacham (2006) I argued that CoC$_M$ fared better than CeC$_E$ and CeC$_L$. All three accounts suffer from continuity worries, but for CoC$_M$ these worries are less severe. And all three accounts are subject to skeptical scenarios—the many brains argument, the sadistic scientists argument or the varied brains argument—but I argued that the varied brains argument and the sadistic scientists argument weren't as damaging as the many brains argument, since they were less likely to apply to agents like us. Given these assessments, CoC_M looks better than its competitors.

My thoughts on these matters have changed, however. As we saw in section 5.10.2, I'm now inclined to think that the varied brains argument is just as damaging as the many brains argument. And I'm now inclined to think that the sadistic scientists arguments—or something like it—is not as damaging as I originally thought. A more detailed explanation of the latter will be provided in chapter 6, but for now, I will just provide a brief sketch of why the sadistic scientists argument doesn't seem as bad to me as the many or varied brains arguments.

First, as we'll see in chapter 7, tricky issues come up when we consider cases involving death, like the sadistic scientists case. In order to sidestep these issues, let's consider a similar case that doesn't involve death:

> *The Narcissistic Scientists Argument:* As before, imagine that you are living in a world where brain-in-a-vat creation technology is cheap and easily accessible. Some narcissistic friends of yours who would enjoy showing off tell you that at midnight they'll spend an hour creating n brains in states subjectively identical to your own. Then, in between 1 am and 2 am, they'll switch the brains from being in a state identical to yours, and instead show the brains various pictures of themselves in glorified poses. Your friends have the resources to carry out this project, and reliably carry out the projects they say they will. If n is big enough, and you hold CeC_L, then though you're now almost certain that they will perform the project, by 2 am you'll be almost certain they didn't. Indeed, if n is big enough, you could even go with them and watch as they create the brains and switch them; if you watch for long enough you won't believe your eyes!

All of the skeptical scenarios I've presented share the following common elements. First, your credence in some unlikely hypothesis about the world will go from very low to very high. Second, this change seems inevitable: no matter what evidence you get, your credence in this hypothesis will go up. The results of these scenarios seem odd because it doesn't seem like your evidence could justify this belief change, given that your beliefs would have changed this way no matter what evidence you got.

We will look at why this seems odd, and when such feelings are justified, in more detail in chapter 6. But for now, let me show that the many and varied brains arguments violate this intuition in a way in which the narcissistic scientists argument does not.

In the narcissistic scientists case your beliefs about the world only change in cases where (i) you are uncertain about what your evidence will be, and (ii) the evidence you might get could push your credence either way. So in the span between midnight and 1 am you know precisely what your evidence will be—you'll appear to be doing what the original is doing, regardless of whether you're the original or a brain-in-a-vat duplicate—and your beliefs about the world don't change. In the span between 1 am and 2 am you are unsure about what evidence you'll get—you might suddenly start seeing portraits of your friends. And in this span, your credence that your friends carried out the project will depend on what evidence you get: if you start seeing portraits, your credence that they carried out the project will go up, if you don't see portraits, your credence that they carried out the project will go down.

This is not the case in the many or varied brains arguments. In the many brains argument you know exactly what your evidence will be, and your credence that there is a scientist creating brains-in-vats will increase anyway. In the varied brains argument you don't know what your evidence will be, but your credence that you're in a strange world will go up regardless of what evidence you get. So the worry that one's evidence couldn't justify one's belief change is much more acute in the many and varied brains cases than it is in the narcissistic scientists case.

Should we conclude from this that CeC$_L$ isn't as bad as CeC$_E$ or CoC$_M$? I would if not for two things. First, the worry I mentioned in 5.9.1: CeC$_L$ looks *ad hoc*. Without a plausible rationale for the asymmetric Continuity Priniciple it employs and the No-Increase Principle, CeC$_L$ looks too arbitrary and unmotivated to be a serious alternative. Second, CeC$_L$ is vulnerable to other skeptical scenarios which are as bad as the many and varied brains arguments. We'll see a scenario of this kind in chapter 7 (section 7.6.4).

Chapter 6
Learning Principles

6.1 Introduction

Each of the three accounts we saw in chapter 5 are susceptible to counterintuitive skeptical scenarios—scenarios that involve inevitable and radical belief change. In the many brains case, the subject becomes virtually certain that brains-in-vats are being created, even though her initial credence in the possibility is low and she knows exactly what her evidence will be. In the narcissistic scientists case, the subject becomes virtually certain that her friends aren't creating brains-in-vats, even though her initial credence in the possibility is high and she knows exactly what her evidence will be. And in the varied brains case, the subject becomes virtually certain that she is in a strange world, even though her initial credence in the possibility is low and she knows that her credence in the possibility will go up regardless of her evidence.

The counterintuitive aspects of these cases might remind one of the Reflection Principles proposed by Van Fraassen (1995). Reflection requires an agent who knows her future credences to have the same credences now. In each of the cases given above, it seems something like Reflection is violated: the subject knows that her future credences will be quite different from her current ones.

This thought is on the right track. But Reflection isn't what we want. For one thing, as a number of authors have pointed out, violations of Reflection aren't generally counterintuitive.[1] (It seems reasonable to be confident about the outcome of the card game you just played, even though you know you won't remember it ten years from now.) So the source of the counterintuitiveness of these scenarios can't just be due to a violation of Reflection. For another, the worry I've raised is a diachronic worry, a worry about how one's credences ought to change over time. But Reflection is a synchronic constraint, imposed on an agent's credences at a single time.[2] So Reflection doesn't directly bear on the kind of worry we're interested in.

In the next section I'll present some principles which I take to capture the fundamental intuitions behind the skeptical arguments. In the section after that, I'll use them to assess these skeptical scenarios.

6.2 The Learning Principles

All of the skeptical scenarios share the following common elements. First, your credence in some implausible hypothesis about the world will greatly increase. Second, this belief change is inevitable: your credence in the hypothesis will go up no matter what evidence you get. These belief changes seem irrational because they don't seem justified by the evidence, given that your beliefs will change in this way regardless of your evidence.

At a first pass, we might try to formulate this intuition as follows. Recall that a subject's doxastic evidence is the evidence her doxastic temporal successors get. Or, restricting ourselves to possibilities in which the subject has a positive credence, the evidence the subject thinks she might get next. Then:

> A rational account of belief updating R must be such that a subject's current credence lies in the span of the credences that R assigns given her doxastic

[1] See, for example, Christensen (1991) and Weisberg (2007*b*).

[2] See Weisberg (2007*a*) for an eloquent discussion of this point.

evidence.

Inevitable belief changes—changes which occur regardless of which bit of doxastic evidence you get—violate this principle, since your current credences won't be the same as, and therefore won't lie in the span of, the credences R will assign you.

But this formulation requires three amendments. First, we only want it to apply to beliefs about what the world is like—*de dicto* beliefs—not to self-locating beliefs. After all, we don't want to reject an account for suggesting that an agent's beliefs about the time will inevitably change as time passes.

Second, we only want the principle to apply when the subject's doxastic temporal successors haven't lost information about what the world is like, i.e., haven't increased their space of doxastic worlds. Consider an agent who knows that a coin toss landed heads, but whose doxastic temporal successors all forget this outcome. We don't want to reject an account which will assign her a credence of $\frac{1}{2}$ in heads, just because her current credence in heads (1) won't lie in this span.

Third, we only want the principle to apply when all of the subject's doxastic temporal successors are around to provide the appropriate credence span. Suppose a subject will be executed in a minute if a coin toss lands tails. If she finds herself alive after a minute, we don't want to reject an account which will assign her a credence of 1 in heads, even though her prior credence in heads ($\frac{1}{2}$) won't lie in this span.

Applying these amendments yields the following principle:

The Learning Principle (LP$_1$): A rational account of updating R must be such that if (i) all of a subject's doxastic alternatives have temporal successors, and (ii) none of these successors suffer from *de dicto* information loss, then her *de dicto* credences will lie in the span of the credences R assigns her given her doxastic evidence.

LP$_1$ applies only to the credences R assigns a subject given the evidence she thinks she might get next—her doxastic evidence. I.e., LP_1 applies only "one step" in advance. But we might also want a constraint of this kind to apply more than one step in advance. For example, we might want to constrain the credences R assigns a subject given the next *two* pieces of evidence she thinks she might get—her doxastic 2-step evidence. Likewise, we might want to constrain the credences R assigns a subject given the next n pieces of evidence she thinks she might get—her doxastic n-step evidence. Extending the principle in this way gives us:

The n-step Learning Principle (LP_n): A rational account of updating R must be such that, for any n, if (i) for any $m < n$, all of a subject's doxastic m-step temporal successors have temporal successors, and (ii) none of a subject's doxastic n-step temporal successors suffer from *de dicto* information loss during those n-steps, then her *de dicto* credences will lie in the span of the credences R assigns given her doxastic n-step evidence.

Why do we need two separate principles? Why doesn't LP_1 entail LP_n? Here's the short answer: given that the antecedent conditions are satisfied, LP_n requires a subject's credences to lie in the span of the credences of her doxastic n-step successors, while LP_1 just requires a subject's credences to lie in the span of the credences of the doxastic successors of the doxastic successors of... (n-times), a superset of her doxastic n-step successors.

To make this more transparent, let's look at an example which shows why LP_1 doesn't entail LP_2. Consider a subject, her successor, and the successor of her successor. Call them 1, 2 and 3, respectively; call their credences $cr1$, $cr2$ and $cr3$; and call their evidence e_1, e_2 and e_3. Further consider a subject in a state subjectively identical to 2's, and her successor. Call this other subject 2^* and her successor 3^*, let $cr2^*$ and $cr3^*$ be their credences, and let e_{2^*} and e_{3^*} be their evidence. (Since 2 and 2^* are subjectively identical, we know that $e_2 = e_{2^*}$).

What constraints does LP_1 impose in this case? LP_1 requires that a subject's current credences lie in the span of the credences R assigns given the evidence of her doxastic temporal successors. 1 knows her temporal successor will be 2, so 2 is her only doxastic temporal successor. So LP_1 requires $cr1$ to lie in span of $cr1^R_{e_2}$ (i.e., requires $cr1 = cr1^R_{e_2}$).

Likewise, 2's doxastic temporal successors are 3 and 3^*. So LP_1 requires $cr2$ to lie in span of $cr2^R_{e_3}$ and $cr2^R_{e_{3^*}}$. Or, equivalently, if subjects update using R, LP_1 requires $cr1^R_{e_2}$ to lie in the span of $cr1^R_{e_2,e_3}$ and $cr1^R_{e_2,e_{3^*}}$. Since "lying in the span of" is transitive, it follows that LP_1 requires $cr1$ to lie in the span of $cr1^R_{e_2,e_3}$ and $cr1^R_{e_2,e_{3^*}}$.

What constraints does LP_2 impose over and above LP_1? LP_2 requires that a subject's credences lie in the span of the credences R would assign given the evidence of her doxastic 2-step temporal successors. Since 1 knows her temporal successor is 2, and that 2's temporal successor is 3, 3 is her only doxastic 2-step temporal successor. So LP_2 requires $cr1$ to lie in the span of $cr1^R_{e_2,e_3}$ (i.e., requires $cr1 = cr1^R_{e_2,e_3}$).

But, as we've seen, LP_1 does *not* require this—it just requires $cr1$ to lie in the span of $cr1^R_{e_2,e_3}$ and $cr1^R_{e_2,e_{3^*}}$. And this is a strictly weaker constraint than LP_2 requries. So LP_1 does not entail LP_2.

Learning Principles and Permissive Accounts

In characterizing the Learning Principles of section 6.2, we implicitly assumed that the accounts in question assign a definite value to every possibility. In chapter 8 we'll look at some "permissive accounts" where this is not the case—accounts which allow the subjects to adopt any credence within some interval. How should we extend the Learning Principles to permissive accounts?

The Learning Principles are supposed to capture the intuitions violated in the skeptical arguments. An account that satisfies the Learning Principles shouldn't allow belief changes that lead to skeptical results. So every credence assignment a permissive account allows must satisfy the $\mathrm{LP}_1/\mathrm{LP}_n$ given in the last section. I.e., one's prior credences must lie in the span of every future assignment that the permissive account allows the subject to adopt.

6.3 Assessing the Skeptical Arguments

Why are the skeptical scenarios we looked at in chapter 5 counterintuitive? The common denominator is that all three violate LP_n.

In all three scenarios, the subjects satisfy the antecedents of LP_n: none of the subjects are destroyed, and none of the subjects suffer from *de dicto* information loss. And in each of the skeptical arguments, if the subject updates using the account in question, then the consequent of LP_n is violated: her current *de dicto* credences won't lie in the span of the credences the account assigns given her doxastic n-step evidence. In the many brains case, the subject's credence that brains-in-vats are being created is very low, but she will become virtually certain that such brains are created if she updates using CeC_E. In the narcissistic scientists case, the subject's credence that her friends aren't creating brains-in-vats is very high, but she will become virtually certain that these brains weren't created if she updates using CeC_L. In the varied brains case, the subject's credence that she is in a strange world is very low, even though she will have a very high credence in the possibility if she updates using CoC_M.

This is why, I think, the results of the skeptical arguments are counterintuitive. LP_n strikes us as a plausible constraint on rational accounts of updating, and the skeptical arguments present us with belief changes that violate this constraint.

But there is more to the story. Let's turn to the other Learning Principle, LP_1.

CeC_E's treatment of the many brains scenario violates this principle as well.

Let t be the time at which the first brain would be created. Let the subject's credence that she is in a brain-creating world right before t be p. Right after t, CeC$_E$ will inevitably assign her a credence of $\frac{2p}{1+p} > p$ that she is in a brain-creating world. So the many brains argument shows that CeC$_E$ violates LP$_1$.

CoC$_M$'s treatment of the varied brains scenario also violates this principle. Let t be the time at which the first batch of brains would be created. Right after t, CoC$_M$ will inevitably assign the subject a higher credence that she is in a strange world then she had before. So the varied brains argument shows that CoC$_M$ violates LP$_1$.

What about CeC$_L$'s treatment of the narcissistic scientists case? In this case, LP$_1$ is not violated (though in chapter 7, we'll see that there are cases in which CeC$_L$ does violate LP$_1$). Consider the first stage of the case, where the brains might be created. During this stage, the subject's credence that the scientists are creating these brains remains constant, so LP$_1$ isn't violated. In the second stage of the case, when the brains (if there are any) would be shown portraits of the scientists, her credence that the scientists are creating the brains will change when she looks around and doesn't see any portraits. This doesn't violate LP$_1$ either, since her prior credence that the brains were created lies in the span of the credences CeC$_L$ assigns given her doxastic evidence: given the "no portraits" evidence CeC$_L$ assigns her a lower credence in the brains having been created, and given the "portraits" evidence CeC$_L$ assigns her a higher credence in the brains having been created.

At the end of chapter 5 I gave some reasons for thinking that the narcissistic scientists result wasn't as bad as the many or varied brains results. With the Learning Principles, we can pinpoint why. While accounts that succumb to the many and varied brains arguments violate both of these principles, accounts that succumb to the narcissistic scientists argument only violate LP$_n$. The Learning

Principles encode certain intuitions we have about when evidence can justify a belief change. And the narcissistic scientists result does less damage to these intuitions than the many or varied brains results.

Chapter 7
Successor Conditionalization

7.1 Introduction

We have reason to be dissatisfied with the three accounts we've looked at so far because they violate the Learning Principles. But our work up to this point has not been without reward. Our investigations have given us a clearer picture of what we want, and what pitfalls to avoid. What we'd like is a well-motivated account which satisfies the Learning Principles.

In this chapter we'll attempt to construct some rules which satisfy the Learning Principles. In the next section we'll look at one attempt to do this, temporal successor conditionalization. In the section that follows, we'll look at some virtues of this account. In the section after that, we'll look at some worries for it. With these worries in mind, we'll turn to look at another account, epistemic successor conditionalization. In the next three sections we'll look at this account, and assess its virtues and vices. In the final section we'll look at how each of these accounts handles death.

7.2 Temporal Successor Conditionalization

We want a rule that satisfies the Learning Principles. Since LP_n entails LP_1, a rule which satisfies LP_n will satisfy both principles. To satisfy LP_n we need a rule on which the current credences of a subject will lie in the span of the credences it assigns given the various pieces of evidence of her doxastic temporal successors. A straightforward way to get such a rule is to directly build in such a requirement.

Here's a natural way of doing this. Call the following rule, *temporal successor conditionalization*:

Temporal Successor Conditionalization (TSC): If a condition-satisfying[1] agent with credences cr gets evidence e, then her new credence in a centered proposition a, $cr_e(a)$, should be:[2]

$$cr_e(a) \overset{\text{def}}{=} \sum_{(i|c_i \in a)} cr(\bar{c}_i) \cdot m_{\bar{c}_i}(c_i e | dts(cr)) \cdot N, \tag{7.2}$$

where N is the normalization factor,

$$N \overset{\text{def}}{=} \sum_j \frac{1}{cr(\bar{c}_j) \cdot m_{\bar{c}_j}(c_j e | dts(cr))}. \tag{7.3}$$

Although the formula expressing TSC is not particularly transparent, it is easy to calculate what a subject's new credences should be according to TSC using diagrams, in the following way:

1. Draw a box, which represents the space of the subject's previous doxastic temporal successors, *dts*.

2. Divide the width of the box into smaller boxes representing the worlds that have members in *dts*, with the width of these boxes proportional to her previous credence in those worlds.

[1] I.e., and agent who satisfies A1, A2, A3*, A4, A5 and A6

[2] As formulated, TSC divides credences equally among the centered worlds at a world. If we dislike this, we can modify the rule to assign credences to alternatives proportional to the ratio of the credence of the world they're in that was assigned to their temporal predecessor. I.e., defining "$Tp(a)$" as the union of the temporal predecessors of the centered worlds of a, we can replace (7.2) with:

$$cr_e(a) \overset{\text{def}}{=} \sum_{(i|c_i \in a)} cr(\bar{c}_i) \cdot \frac{cr(Tp(c_i e))}{\sum_{(i|c_i \in \bar{c}_i edts(cr))} cr(Tp(c_i))} \cdot N, \tag{7.1}$$

where N is the appropriate normalization factor. This amendment complicates things in various ways, however, so in what follows, I'll stick with the simpler formulation given above.

3. Divide the height of each of the smaller boxes into boxes representing the doxastic temporal successors at that world, with the height divided equally.

4. Eliminate every box incompatible with the subject's new evidence.

5. The subject's new credences in these possibilities are proportional to the area of the box that represents them.

Let's apply this to the sleeping beauty case. On Sunday night, right before she goes to sleep, Beauty's doxastic temporal successors consist of her Monday morning temporal successors at the heads worlds, and her Monday morning temporal successors at the tails worlds, with her credence divided evenly between them. So we can draw the space of her doxastic temporal successors like this (where I've put the evidence compatible with each possibility in parentheses):

MON(e)	MON(e)
H	T

When she wakes up, her evidence e is compatible with all of these alternatives, so none of them are eliminated. Since her new credences in these possibilities are proportional to their area, we can conclude that her new credences according to TSC should be $cr_e(\text{H}\wedge\text{MON}) = cr_e(\text{T}\wedge\text{MON}) = \frac{1}{2}$. What about her credence in T∧TUE? On Sunday night her Tuesday morning continuants are not doxastic temporal successors, so her credence in this possibility should be 0.

If Beauty adopts these credences, then she is doing well as far as correctness goes: she is certain that it's Monday on Monday. But it's reasonable to wonder whether it's fair to expect her to adopt these credences. After all, when she wakes up on Monday morning she doesn't know who her doxastic temporal successors were, and thus what credences TSC assigns her. To put it another way, one might reasonably wonder how useful TSC is as a source of guidance for a subject like Beauty. We'll return to these worries in section 7.4.

What should Beauty's credences be if, shortly after she wakes up, she will be told that it's Monday? Well, her doxastic temporal successors before she is told should consist of her Monday morning temporal successors at the heads and tails worlds, with her credence divided evenly between them. (Her Tuesday morning continuant should not be a doxastic temporal successor if she's following TSC, since TSC assigned it a credence of 0.) So her *dts* should be:

MON("MON")	MON("MON")
H	T

Since her evidence that it's Monday is compatible with both of these possibilities, her credences should remain the same: $cr_{``MON''}(\text{H}\wedge\text{MON}) = cr_{``MON''}(\text{T}\wedge\text{MON})$ $= \frac{1}{2}$.

7.3 Virtues of TSC

7.3.1 Generic Virtues of TSC

TSC shares the generic virtues of the other accounts we've looked at so far. TSC satisfies the probability axioms (see Appendix 9.16 for the proof). And in standard conditions, TSC gives the same results as classical Bayesianism. The proof to the latter claim is given in Appendix 9.17, but the intuitive idea is clear enough. In standard conditions one has only one temporal successor at each doxastic world, and the objects and evidence are *de dicto* propositions. In these circumstances, the procedure for calculating your new credences given above will reduce to this: take your current credences in worlds, eliminate those worlds incompatible with your evidence, and reassign credences to the survivors such that the proportions between their credences remains the same. And this is just the classical Bayesian rule.

7.3.2 TSC Satisfies the Learning Principles

The main virtue of TSC is that it satisfies both Learning Principles. The proof is provided in Appendix 9.18, but we can get a feel for why this is the case be examining how TSC treats each of the three skeptical scenarios we looked at in chapter 5.

Let's start with the many brains case. Consider your *dts* right before the first brain would be created. Your *dts* is the union of your temporal successors at all of your doxastic worlds. At worlds where brains are not being created, you will have a single temporal successor. At worlds where brains are being created, you will also have a single temporal successor, since you know that the brain that's about to be created isn't one of your temporal successors. So your *dts* will consist of only your temporal successor at each of your doxastic worlds. And since your evidence doesn't eliminate any of these successors, your credence in whether you're in a brain-creating world will remain the same.

To see this with diagrams, let S be the hypothesis that there is a scientist who will be creating brains in the specified manner, and let e be the evidence that your successor will receive. Then the diagram of your *dts* right before the first brain is created will be:

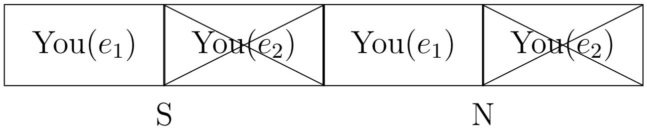
You(e)	You(e)
S	¬S

Since the evidence you get, e, won't eliminate any of these possibilities, your credences in S and ¬S will remain the same.

The same will be true when the next brain would be created, and the brain after that. So your credences will remain stable throughout. Since your credences will lie in the span of the credences TSC assigns you given your doxastic n-step evidence, LP$_n$ and LP$_1$ will be satisfied.

Next, let's look at the varied brains case. As with the many brains case, your *dts* right before the first brains would be created will contain only your temporal successor at each world. Since you have a temporal successor compatible with each of the n subjectively distinguishable experiences you might have in the next time step at both the strange worlds and the normal worlds, successors at both the strange and the normal worlds will be eliminated when you get your evidence. Assuming each experience takes up the same proportion of strange worlds and normal worlds, then your credence in being in a strange world will remain the same.

To see this with diagrams, let S be the hypothesis that you live in a strange world where brains are being created in the specified manner, and let N the hypothesis that you are living in a normal world. For simplicity, assume that there are only 2 subjectively distinguishable experiences you might have in the next time step, e_1 and e_2. Then the diagram of your *dts* right before the first brains would be created will be:



Your evidence, e_1 say, will eliminate the same proportion of S possibilities and N possibilities, so your credences in S and N will remain the same.

The same will be true for the next evidence you get, and the evidence after that. So your credences will remain stable throughout. Since your credences will lie in the span of the credences TSC assigns you given your doxastic n-step evidence, LP$_n$ and LP$_1$ will be satisfied.

Finally, let's look at the narcissistic scientists case. The first part of the narcissistic scientists case, where the brains might be created, is essentially identical to the many brains case. And as we've seen, in that case your credence that brains are being created will remain the same. In the second part of the narcissistic scientists case, the brains, if created, are shown portraits of the scientists. But your *dts* right before they would be shown the portraits will consist only of your temporal successor at each world. And since none of your temporal successors will see portraits, your evidence—not seeing a portrait—won't eliminate any of them. So again, your credences will again remain the same.

To see the latter point with diagrams, let S be the hypothesis that the scientists will carry out the project of creating the brains in the specified manner, and let $\neg p$ be the not-seeing-portraits evidence. Then the diagram of your *dts* right before the brains would be shown portraits will be:

You($\neg p$)	You($\neg p$)
S	$\neg$S

Since the evidence you get, $\neg p$, won't eliminate any of these possibilities, your credences in S and $\neg$S will remain the same.

So p will never be part of your doxastic n-step evidence if you update in accordance with TSC, since it insists that your credence that you're a brain-in-a-vat be 0. And in this case your credences will lie in the span of the credences TSC assigns you given your doxastic n-step evidence, since your credences will remain the same throughout. So LP$_n$ and LP$_1$ will be satisfied.

7.4 Worries about TSC

7.4.1 Inegalitarianism: Time Location, Duplication and Fission

Recall the hypothetical prior accounts of chapter 5. In the sleeping beauty case, when Beauty woke up on Monday, they assigned her the same credences regardless of whether the source of her *de se* uncertainty was the possibility of future memory loss, the possibility of duplication or the possibility of fission.[3] So there's a sense in which these accounts are egalitarian. TSC is not egalitarian in this sense. To see this, let's compare how TSC treats three versions of the sleeping beauty case: the original case, a duplication variant and a fission variant.

We saw how TSC treats the temporal uncertainty version of the sleeping beauty case in section 7.2. When Beauty wakes up on Monday morning, TSC assigns her a credence of $\frac{1}{2}/\frac{1}{2}$ to heads and tails, and a credence of 1 to Monday. And if Beauty is then told what day it is, TSC continues to assign her a credence of $\frac{1}{2}/\frac{1}{2}$ to heads and tails, and a credence of 1 to Monday.

How does TSC treat a duplication version of sleeping beauty, where instead of being put to sleep again if the coin lands tails, a duplicate of her is created in a subjectively indistinguishable situation? On Sunday night she knows that her temporal successor will not be a duplicate, so her *dts* will consist of her Monday morning successor at both the heads and the tails worlds:

Original(e)	Original(e)
H	T

Since her waking evidence e doesn't eliminate either of these possibilities, TSC will assign her a credence of $\frac{1}{2}/\frac{1}{2}$ to heads and tails, and a credence of 1 to her being the original, not the duplicate.

What if, a minute after waking, she is told that she's the original? Her *dts* right before being told this will be:

Original("Original")	Original("Original")
H	T

[3] Though technically this needn't be the case: they could treat these cases differently by adopting a Continuity Principle which assigned priors which discriminated between these different possibilities.

Since her evidence doesn't eliminate any of these possibilities, her credence will remain $\frac{1}{2}/\frac{1}{2}$ in heads and tails, and remain 1 in her being the original.

How does TSC treat a fission version of the sleeping beauty case? To make things concrete, let this case be set up in the following way. As before, Beauty will be put to sleep on Sunday night and woken up on Monday morning. And that night, while she is sleeping, the experimenters will flip a fair coin. If the coin lands heads, Beauty will be placed in one of a pair of rooms, the left one, and woken up as usual. If the coin lands tails, then Beauty will be fissioned into two "fissiles", and the first fissile will be placed in the left room, and the second fissile will be placed in the right room.

According to TSC, what should Beauty's credences be when she wakes up? On Sunday night her *dts* consists of her lone successor at the heads worlds, and a pair of successors, the two fissiles, at the tails worlds:

Original(e)	Fissile 2(e)
	Fissile 1(e)
H	T

When Beauty wakes up her evidence, e, won't eliminate any of these possibilities, so her credences in heads and tails will remain $\frac{1}{2}/\frac{1}{2}$, with her credence in tails being split evenly between fissile 1 and fissile 2.

What if, a minute after waking, she is told she's in the left room? Her *dts* right before being told this will again consist of her lone successor at the heads worlds and the pair of fissiles at the tails worlds:

Original("Left")	~~Fissile 2("Right")~~
	Fissile 1("Left")
H	T

Her evidence will eliminate the possibility of her being the second fissile at the tails worlds. Assigning credences to the remaining possibilities in proportion to their area, we find that her credence in heads and tails will become $\frac{2}{3}/\frac{1}{3}$.

So TSC is inegalitarian: it doesn't treat cases of fission in the same way as it treats cases of temporal location or duplication. There are two worries one might raise in light of this unequal treatment.

The first worry concerns the role of personal identity in TSC. TSC is inegalitarian because the manner in which it assigns credences relies on the notion of temporal continuity, and the facts about temporal continuity are not the same in cases of fission and cases of duplication. As we saw in chapter 4, the notion of temporal continuity we're working with is tied to the folk notion of personal identity. (And it is important that temporal continuity be tied to our ordinary notion of personal identity in this way: it is this relation that allowed us in chapter 6 to translate intuitions about our future credences into intuitions about the credences of our temporal continuants.) So another way to see the matter is this: TSC is inegalitarian because the manner in which it assigns credences relies on the notion of personal identity, and the facts about personal identity are not the same in cases of fission and cases of duplication.

This might strike one as an unhappy state of affairs. One might think that a satisfactory updating rule shouldn't rely on something as vague and slippery as the notion of personal identity. And since TSC does rely on such a notion, one might take this to be a reason to be skeptical of TSC.

Although I am sympathetic to the unease this worry evokes, it is not a tenable criticism. It is not just TSC that relies on a notion of personal identity: all updating rules rely on a notion of personal identity. Classical Bayesianism, for example, assigns credences to a subject in light of her evidence and previous credences. This prescription presupposes a way of picking out the same subject at two different times. Likewise, the hypothetical prior accounts we looked at in chapter 5 rely on a notion of personal identity. These accounts rely on the subject's hypothetical priors, and hypothetical priors, however we make sense of them, are required to be static: a subject must have the same hypothetical priors at every time. Again, this presupposes a way of picking out the same subject at different times.

More generally, an updating rule is a diachronic credence constraint: a constraint on how our credences at different times should be related. Any rule of this kind requires a way of picking out a subject at different times. I.e., any rule of this kind requires a notion of personal identity.

The second worry is more straightforward. One might have the intuition that cases of duplication and fission are identical in all epistemically relevant respects. If so, then it seems an updating rule should treat them in the same way. Unlike the first worry, this is a tenable criticism. We'll look at a way to modify TSC in light of this criticism in section 7.5.

Fission and the Narcissistic Scientists

We saw in the last section that TSC treats cases of fission differently from cases of duplication or temporal location. For example, TSC treats the fission version of the sleeping beauty case in the same way that CeC_L does, while its treatment of the duplication version of sleeping beauty diverges from CeC_L.

In section 7.3.2 we saw that TSC avoids the LP_n violation that CeC_L runs into in the narcissistic scientists case. But the narcissistic scientists case is a case of duplication. How does TSC treat a fission variant of the narcissistic scientists case? It seems TSC would treat it in the same way as CeC_L does. But then why doesn't TSC violate LP_n?

Here is the fission version of the narcissistic scientists case:

> Imagine that you are living in a world where fission technology is cheap and easily accessible. Some narcissistic friends of yours who would enjoy showing off tell you that at midnight they'll fission you into n fissiles. They'll put the first fissile back in your original place, and put each of the 2nd through nth fissiles in an environment subjectively identical to your own. Then,

> at 1 am, they'll place portraits of themselves in various glorified poses in the environments of the 2nd through nth fissiles. Your friends have the resources to carry out this project, and reliably carry out the projects they say they will.

What should your credences be at midnight, according to TSC? Your *dts* right before midnight will consist of a single successor at the worlds where the scientists don't perform the fissioning, and of n successors—the n fissiles—at the worlds where the scientists do perform the fissioning. So, letting S stand for the hypothesis that the scientists will carry out the project as promised, your *dts* will be this:

Fissile 1(e)	Fissile 2(e)	Fissile 3(e)	···	Fissile n(e)	You(e)
		S			¬S

Your evidence at midnight, e, won't eliminate any of these possibilities. So according to TSC, your credence in S and ¬S should remain the same, with your credence in S divided evenly between the n fissiles.

What should your credences be at 1 am if you don't see any portraits? Your *dts* right before 1 am will again consist of a single successor at the worlds where the fissioning isn't performed, and of the n fissiles at the worlds where fissioning is performed. Letting p stand for the evidence you'll get if you see a portrait, your *dts* right before 1 am will be:

Fissile 1($\neg p$)	~~Fissile 2(p)~~	~~Fissile 3(p)~~	···	~~Fissile n(p)~~	You($\neg p$)
		S			¬S

At 1 am your evidence, $\neg p$, will eliminate most of the possibilities in S, and none of your possibilities in ¬S. If your credences are proportional to the diagram given above, your new credences in S and ¬S will be $\frac{1}{2}/\frac{1}{2}$. If your credences aren't proportional to the dimensions shown in the diagram, and $cr(\neg S) \gg \frac{cr(S)}{n+1}$ (where

cr is your credence right before 1 am), then you will become virtually certain that the scientists did not perform the fissioning.

These are the same credences that CeC$_L$ assigns to this case. In the duplication version of the narcissistic scientists case these credence assignments violate LP$_n$. So why don't they violate LP$_n$ in this case too?

CeC$_L$ violates LP$_n$ in the duplication version of this case because your initial credence in S is high, and yet, given any of your doxastic n-step evidence, CeC$_L$ assigns you a very low credence in S, $x \ll 1$. Of course, if you had doxastic n-step evidence that included p, you'd be fine—you'd have some doxastic n-step evidence on which CeC$_L$ would assign you a credence of 1 in S, and thus your initial credence in S would lie in the span of the credences that CeC$_L$ would assign you given your doxastic n-step evidence: $cr(S) \in [x, 1]$. But since you know at the start that you're not a brain-in-a-vat, you know that p is not evidence that *you* will ever get—it's not part of your doxastic n-step evidence—so CeC$_L$ violates LP$_n$.

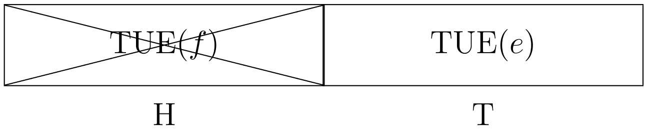
In the fission version of this case, the brains-in-vats are replaced by fissiles. And unlike the brains-in-vats, the evidence that the fissiles get is evidence you think *you* might get; i.e., a part of your doxastic n-step evidence. Since many of these fissiles will get evidence p—evidence which eliminates all of the ¬S worlds—you will have some doxastic n-step evidence on which CeC$_L$ assigns you $cr(S) = 1$. So in the fission version of this case, your initial credences *will* lie in the span of the credences CeC$_L$ assigns given your doxastic n-step evidence, and LP$_n$ is satisfied.

So it's only the duplication version of the narcissistic scientists case, not the fission version, that leads to LP$_n$ violations for CeC$_L$. And since TSC only agrees with CeC$_L$ in the fission case, TSC doesn't violate LP$_n$.

7.4.2 Guidance

The biggest worry for TSC is that it often won't provide useful guidance.

Consider the credences TSC assigns Beauty in the sleeping beauty case. As we saw in section 7.2, when Beauty wakes up on Monday, her credence in heads and tails should be $\frac{1}{2}/\frac{1}{2}$ according to TSC, and her credence that it is Monday should be 1. What should her credences be, if the coin lands tails, when she wakes up on Tuesday? According to TSC, her *dts* right before she is put to sleep on Monday should consist of a single successor at both the heads and the tails worlds, with her credence evenly divided between them:



When she wakes up on Tuesday her evidence will be the same as when she woke up on Monday: e. This will eliminate the successor at the heads world, and so her credence in heads and tails according to TSC should be 0/1, and her credence that it is Tuesday should be 1 as well.

Of course, when Beauty wakes up she doesn't know whether it's Monday or Tuesday, so she doesn't know what her prior *dts* and credences were. If it's Monday morning her prior credences were cr(H) = cr(T) = $\frac{1}{2}$ and cr(SUN) = 1; if it's Tuesday morning her prior credences were cr(H) = cr(T) = $\frac{1}{2}$ and cr(MON) = 1. TSC provides *some* guidance in this case—the subject knows that her credence in heads should be between 0 and $\frac{1}{2}$, and that her credence in tails should be between $\frac{1}{2}$ and 1. But this is probably not as much guidance as Beauty would like.

Another lapse of guidance comes in cases of duplication. Consider the credences TSC assigns Beauty in the duplication version of the sleeping beauty case. As we saw in section 7.4.1, when Beauty wakes up her credence in heads and tails should be $\frac{1}{2}/\frac{1}{2}$, and her credence that she is the original should be 1. But what should the credences of her duplicate be when the duplicate wakes up? The duplicate has no temporal predecessor, and thus has no *dts*. So TSC doesn't place any constraints on her credences. In this case, TSC provides even less guidance than in the original sleeping beauty case. Since Beauty doesn't have access to information that will tell her whether she's the duplicate, all she knows is that her credences should be cr(H), cr(T), cr(Original)$\in [0, 1]$.

Although TSC comes up short as a guidance tool in the original and duplication versions of the sleeping beauty case, it provides good guidance in the fission version of the sleeping beauty case. As we saw in section 7.4.1, when Beauty wakes up on Monday her credences according to TSC should be cr(H) $=$ cr(T) $=$ $\frac{1}{2}$ and cr(fissile 1) $=$ cr(fissile 2) $= \frac{1}{4}$. And since both fissiles will have predecessors with the same credences and *dts*, TSC will assign the fissiles the same credences when they wake up.

So the news with respect to guidance isn't all bad: in most cases of fission, TSC provides subjects with perfect guidance. But this isn't much of a consolation for subjects stuck in the original sleeping beauty case.

Here is the source of TSC's guidance failures. TSC generates a subject's new credences from her evidence and her previous credences. In cases where a TSC-following subject doesn't have access to her previous credences, she can fail to know what credences to adopt. This is what happens in the original and duplication versions of the sleeping beauty case. In the fission version of the sleeping beauty case, on the other hand, the subject does know what her previous credences were, and so TSC does provide guidance.

To put this in context, note that classical Bayesianism fails to provide guidance in precisely the same way. Like TSC, classical Bayesianism generates a subject's new credences from her evidence and her previous credences. And in cases where subjects don't have access to their previous credences, classical Bayesianism fails to provide useful guidance. For example, consider a standard case of memory loss, where a subject's temporal predecessor knows the outcome of a coin toss, but the subject has forgotten what it was. Classical Bayesianism assigns this subject her predecessor's credence in heads. But this won't be helpful guidance-wise, since the subject doesn't have access to her predecessor's credences.[4]

So how should we assess this worry? If we're looking for an updating rule which provides guidance in cases like the original sleeping beauty case, we'll be unhappy with TSC. We'll return to discuss the extent to which this expectation is reasonable, and the limits of guidance, in chapter 8.

[4] In section 5.1 we saw two reasons why classical Bayesianism can't handle self-locating beliefs: a superficial reason—it doesn't work with possibilities that are fine-grained enough to capture self-locating beliefs—and a deep reason—it doesn't allow for the addition as well as the elimination of possibilities. And we saw that it's the deep reason that's the source of the real trouble. Arntzenius (2003*a*) has pointed out that a similar problem for classical Bayesianism arises in cases of *de dicto* information loss. For example, suppose a subject sees a coin land heads, and then forgets this outcome. Classical Bayesianism will still require the subject to have a credence of 1 in heads after they've forgotten the outcome, since it doesn't allow her to expand her doxastic worlds by adding the tails worlds back in. And it seems unreasonable to expect the subject to have the same credence in heads as before, since she has no way of knowing whether her prior credence in heads was 0 or 1. As in the self-locating case, the problem is that classical Bayesianism doesn't allow for the addition as well as the elimination of possibilities.

The criticism Arntzenius (2003*a*) raises is essentially that classical Bayesianism fails to provide guidance for subjects in cases of *de dicto* information loss. Given this, it's natural to wonder whether the deep reason that classical Bayesianism can't handle self-locating beliefs is, at bottom, a matter of guidance as well.

Although both criticisms focus on the same feature of classical Bayesianism, they are different kinds of criticisms. The fact that classical Bayesianism doesn't allow for the addition of possibilities means that a subject who is certain that it is 9 am at 9 am must remain certain that it's 9 am at every time after that. This makes it impossible for a subject to always have correct beliefs. And there is something wrong with a rule which requires maximally informed agents—such as God, if she exists—to have incorrect beliefs.

The problem in the case of *de dicto* information loss is not that the subject's credences aren't correct. Indeed, if she adopts the credences that the rule suggests, then her credences will be more correct than if she does not. The problem is that it seems unreasonable to expect a subject to adopt the credences classical Bayesianism assigns her in cases of *de dicto* information loss.

So though these two problems arise from the same feature of classical Bayesianism—the fact that it doesn't allow for the addition of possibilities—they are different kinds of criticisms. One criticizes the correctness of the credences classical Bayesianism would assign, the other criticizes the classical Bayesianism's inability to provide guidance.

7.5 Epistemic Successor Conditionalization

TSC was a mixed success. On the one hand, it succeeded in satisfying the Learning Principles, which is what we constructed it to do. On the other hand, it's inegalitarian: it treats cases where one's self-locating uncertainty is due to duplication, to temporal location and to fission in different ways. And in a number of self-locating cases, it fails to be a useful guidance tool.

How might we modify TSC to get around the first worry? TSC treats cases of fission, duplication and temporal uncertainty differently because how it assigns credences depends on the temporal continuity facts, and the temporal continuity facts in fission cases are different than they are in duplication cases or temporal uncertainty cases. But while fission, duplication and temporal uncertainty cases differ with respect to their temporal continuity facts, they do not differ with respect to their epistemic continuity facts. So we can avoid TSC's inegalitarianism by replacing the doxastic temporal successors with doxastic epistemic successors (des).[5] As we will see in section 7.6.3, this modification also works to mitigate the second worry. Call this *epistemic successor conditionalization*:

Epistemic Successor Conditionalization (ESC): If a condition-satisfying[6] agent with credences cr gets evidence e, then her new credence in a centered proposition a, $cr_e(a)$, should be:

$$cr_e(a) \overset{\text{def}}{=} \sum_{(i|c_i \in a)} cr(\bar{c}_i) \cdot m_{\bar{c}_i}(c_i e | des(cr)) \cdot N, \tag{7.4}$$

[5] Though note that such an account may still be inegalitarian in other senses. The credences this account will assign to the duplicate in the duplication version of the sleeping beauty case need not be the same as the credences the account assigns the second fissile in the fission version of the sleeping beauty case, for example.

[6] I.e., and agent who satisfies A1, A2, A3*, A4, A5 and A6

where N is the normalization factor,

$$N \overset{\text{def}}{=} \sum_j \frac{1}{cr(\bar{c}_j) \cdot m_{\bar{c}_j}(c_j e | des(cr))}. \tag{7.5}$$

As with TSC, it is easy to calculate what a subject's new credences should be according to ESC using diagrams. The procedure is identical to that employed with TSC, except that we replace doxastic temporal successors with doxastic epistemic successors:

1. Draw a box, which represents the space of the subject's previous doxastic epistemic successors, *des*.

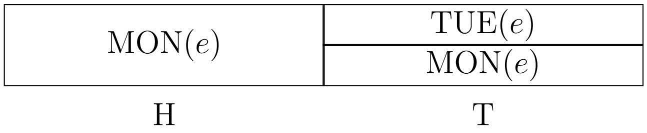
2. Divide the width of the box into smaller boxes representing the worlds that have members in *des*, with the width of these boxes proportional to her previous credence in those worlds.

3. Divide the height of each of the smaller boxes into boxes representing the doxastic epistemic successors at that world, with the height divided equally.

4. Eliminate every box incompatible with the subject's new evidence.

5. The subject's new credences in these possibilities are proportional to the area of the box that represents them.

Let's apply this to the sleeping beauty case. On Sunday night, right before she goes to sleep, Beauty's doxastic epistemic successors will consist of a Monday morning epistemic successor at both the heads and tails worlds, and a Tuesday morning epistemic successor at the tails world. We can draw the space of her doxastic epistemic successors like this:



When she wakes up, her evidence e is compatible with all of these possibilities, so none of them are eliminated. Since her new credences in these possibilities are proportional to their area, we can conclude that her new credences should be $cr_e(\text{H}) = cr_e(\text{T}) = \frac{1}{2}$, with $cr_e(\text{T}\wedge\text{MON}) = cr_e(\text{T}\wedge\text{TUE}) = \frac{1}{4}$.

What should Beauty's credences be if she is then told that it's Monday? Her doxastic epistemic successors before she is told what day it is will consist of her Monday morning epistemic successor at the heads and tails worlds, and her Tuesday morning epistemic successor at the tails world, with her credence divided evenly between the heads and tails worlds. So her *des* will be:

<table>
<tr><td rowspan="2">MON(“MON”)</td><td>~~TUE(“TUE”)~~</td></tr>
<tr><td>MON(“MON”)</td></tr>
<tr><td>H</td><td>T</td></tr>
</table>

Since her evidence is incompatible with the tails and Tuesday possibility, she eliminates it, and assigns new credences to the remaining possibilities in proportion to their area: $cr_{``MON''}(\text{H}\wedge\text{MON}) = \frac{2}{3}$, $cr_{``MON''}(\text{T}\wedge\text{MON}) = \frac{1}{3}$.

Now let's turn to assess the virtues and vices of ESC.

7.6 The Virtues and Vices of ESC

7.6.1 Generic Virtues of ESC

ESC shares the generic virtues of the other accounts we've looked at so far. First, ESC satisfies the probability axioms. To see this, note that: (i) ESC is structurally identical to TSC, (ii) TSC satisfies the probability axioms (see Appendix 9.16), and (iii) these proofs don't make use of any particular facts about *dts* other than that *dts* is a centered proposition.

Second, in standard cases ESC yields the same results as classical Bayesianism. To see this, note that: (i) TSC agrees with classical Bayesianism in standard cases (Appendix 9.17), (ii) ESC yields the same assignments as TSC in cases where all of one's doxastic epistemic successors are doxastic temporal successors, and (iii) in standard cases, all of one's doxastic epistemic successors are doxastic temporal successors.

7.6.2 ESC Satisfies LP_1

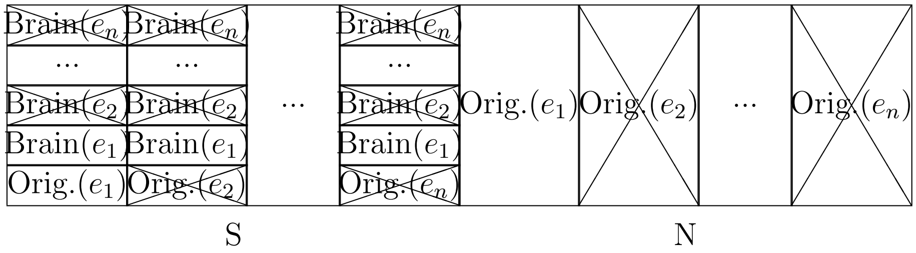
ESC satisfies LP_1 (see Appendix 9.19), but does not satisfy LP_n. To get a feel for why this is the case, let's look at how ESC treats the three skeptical scenarios.

Let's first look at the many brains case. Your *dts* right before the first brain is created will consist of your temporal successor at all of your doxastic worlds. Your *des* consists of your *dts* and any centered worlds at your doxastic worlds that are in the same subjective state as members of your *dts*. So your *des* right before the first brain would be created will consist of your temporal successor at all of your doxastic worlds, and an epistemic successor—the brain-in-a-vat—at the worlds where the scientist is creating brains. Letting S be the hypothesis that there is such a scientist, and letting e be the evidence you'll receive at the time the first brain would be created, we can diagram your *des* like this:

Brain(e)	Original(e)
Original(e)	
S	¬S

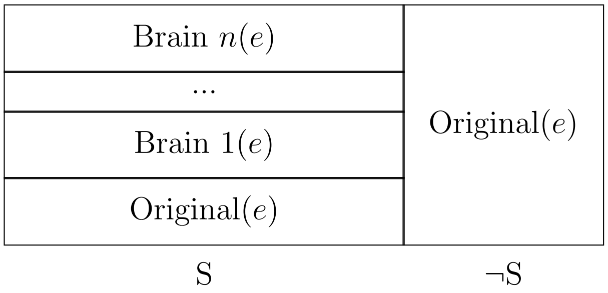
Your evidence e won't eliminate any of these possibilities, so your credence in S and ¬S remains the same. The same is true for the brains that follow, so your credence in S remains constant. So, like TSC, ESC satisfies the Learning Principles in this case.

Let's look at the varied brains case next. Let e_1-e_n be the kinds of evidence you believe you might get. As before, your *dts* right before the first brains are created contains only your temporal successor at each world. But at each of the strange worlds there is also be a brain-in-a-vat in the same subjective state as each of your normal-world successors. So your *des* contains your temporal successors at every world, and an additional n brain-in-vats at each of the strange worlds. Let N be the hypothesis that you're in a normal world, and let S be the hypothesis that you're in a strange world. Suppose that your initial credence in N and S is divided equally among these possibilities, and suppose that the evidence you get is e_1. Your *des* is:

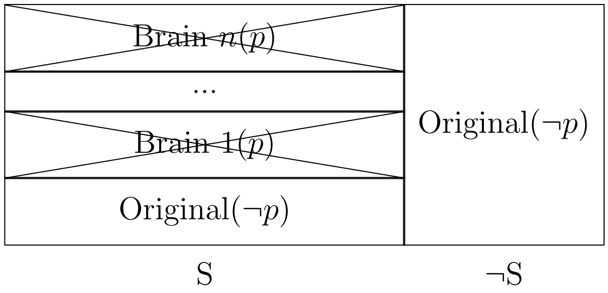


Your evidence e_1 eliminates $\frac{n-1}{n}$ of your N worlds, $\frac{n-1}{n}$ of the brains at each S world, and $\frac{n-1}{n}$ of the originals at your S worlds. Since this eliminates the same proportion of possibilities from both the strange and the normal worlds, your credence in N and S remains the same. And the same would be the case if you got any other evidence. So ESC, like TSC, will satisfies the Learning Principles in this case.

In the previous two skeptical scenarios, ESC satisfied both Learning Principles. But this is not the case in the narcissistic scientists scenario. The first part of the narcissistic scientists case is the same as the many brains case, except that you start with a high credence that brains are being created (S) instead of a low one. Your *des* right before the brains would be created is:

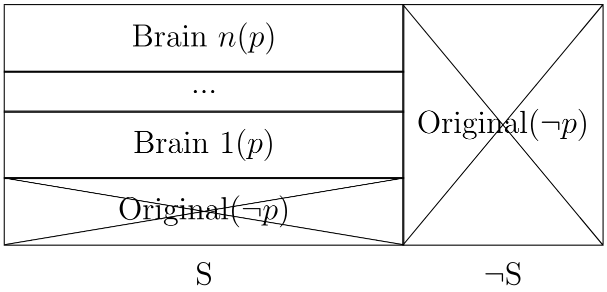


Since your evidence e won't eliminate any of these possibilities, your credence in S remains the same. What happens when the brains, if there are any, are shown the portraits? Your *dts* right before the brains would be shown the portraits consists of the temporal successor of the original at every doxastic world, and the temporal successor of every brain you think you might be at the brain-creating worlds. (Why weren't these part of your *dts* on TSC? Because unlike ESC, TSC assigns you a 0 credence in being one of the brains, and thus prevents them from becoming part of your *dts*.) Since there won't be any other centered worlds in your doxastic worlds that are in the same subjective state as members of your *dts*, your *des* is identical to your *dts*. So your *des* is:



Since you won't see any portraits, your evidence $\neg p$ eliminates all of the brain possibilities, and your credence in ¬S increases. If $cr(\neg S) \gg \frac{cr(S)}{n+1}$, you'll become virtually certain that the scientists did not create the brains. So ESC treats the narcissistic scientists case just like like CeC$_L$ does, and so, like CeC$_L$, violates LP$_n$.

However, like CeC$_L$, ESC does not violate LP$_1$ in this case. On ESC you come to have a non-zero credence that you're a brain-in-a-vat, and so the evidence that the brains would get when the portraits are shown—p—counts as part of your doxastic evidence. ESC would assign you a credence of 1 in S if you got evidence p:



So your prior credence in S will lie in the span of the credences ESC assigns given your doxastic evidence, and LP_1 will be satisfied.

7.6.3 Guidance

Structurally, ESC is similar to TSC and classical Bayesianism. Like those rules, it generates a subject's new credences from her evidence and her previous credences. Since the guidance problems TSC faces stem from the fact that it is a rule of this kind, it's inevitable that ESC will fail to provide guidance in many of the same kinds of cases. But ESC does marginally better guidance-wise than TSC does.

TSC generates a subject's new credences from her evidence and her previous credences. The aspects of the subject's previous credences that it employs are her previous *de dicto* credences and her previous *dts*. TSC runs into guidance problems in cases where the subject doesn't have access to the facts about her prior credences that she needs to know in order to determine the credences TSC assigns. But we can make it such that there are fewer guidance problems if we make what you need to know about your prior credences less descriminating—less likely to vary from person to person. And a subject's *des* is less discriminating than her *dts*. (We'll see some examples of how this works, below.)

Let's start by looking at some cases where ESC fails to provide guidance. Like TSC, ESC provides little guidance in cases involving newly created duplicates. This is because ESC constrains how a subject's current credences relate to her prior ones, and newly created duplicates don't have prior credences.[7]

And like TSC, ESC fails to provide guidance in cases of memory loss or *de dicto* information loss which leave the subject without access to crucial parts of her prior beliefs. For example, consider again the case where a subject's temporal predecessor knows the outcome of a coin toss, but the subject has forgotten what it was. ESC assigns this subject the same credence in the outcome as her predecessor. But since the subject doesn't have access to her prior credences, this advice isn't helpful. So ESC fails to provide guidance in many of the same cases as TSC.

ESC and TSC are also alike with respect to their treatment of the duplication and fission variants of the sleeping beauty case. Both fail to provide guidance in the duplication version of the sleeping beauty case, since the duplicate doesn't have prior credences to constrain her new ones. Both provide perfect guidance in the fission version of the sleeping beauty case. Since Beauty's *des* is the same as her *dts* in this case, ESC and TSC provide identical prescriptions.

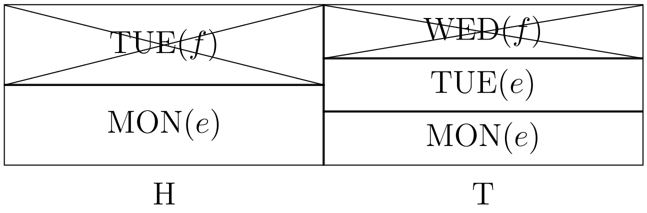
Now let's turn to look at the original sleeping beauty case. TSC fails to provide guidance in this case, and ESC has similar problems, though to a smaller degree. As we saw in section 7.5, when Beauty wakes up on Monday ESC assigns her $\mathrm{cr(H)} = \mathrm{cr(T)} = \frac{1}{2}$ and $\mathrm{cr(T{\wedge}MON)} = \mathrm{cr(T{\wedge}TUE)} = \frac{1}{4}$.

What about when Beauty wakes up on Tuesday in the tails worlds? Let's look at her prior *des*, right before she goes to sleep on Monday night. On Monday night she thinks one of three possibilities obtains: H∧MON, T∧MON or T∧TUE. If T∧MON obtains, then her temporal successor will wake up on Tuesday morning with no memory of having woken before: call this evidence e. If T∧TUE obtains, then her temporal successor will wake up on Wednesday, with the memory of having woken up before: call this evidence f. And if H∧MON obtains, then her

[7] In chapter 8, though, we'll look at some natural ways of extending ESC to accommodate cases of this kind.

temporal successor will wake up on Tuesday morning, with the memory of having woken up before: evidence f.

All of one's temporal successors are epistemic successors, so these three successors will be part of her *des*. Her *des* also includes any centered worlds at her doxastic worlds that are in the same subjective state as her temporal successors. Since she also gets evidence e when she wakes up on Monday—she will wake up with no memory of having woken before—her *des* will include these centered worlds. So her *des* right before she wakes up will be:



Her evidence e will eliminate one of the possibilities at each of these worlds. Assigning credences to the survivors in proportion to their area, we find that her credences according to ESC should be cr(H) = $\frac{3}{7}$ and cr(T) = $\frac{4}{7}$. But since Beauty doesn't know whether it's Monday or Tuesday when she wakes up, she doesn't know whether her credences should be $\frac{1}{2}/\frac{1}{2}$ or $\frac{3}{7}/\frac{4}{7}$. So although these credences are closer together than the ones TSC assigns—$\frac{1}{2}/\frac{1}{2}$ and 0/1—ESC fails to provide guidance.

Next consider the version of the sleeping beauty case in which Beauty is told what day it is shortly after waking. We've already seen that Beauty's credences after waking on Monday are cr(H) = cr(T) = $\frac{1}{2}$ and cr(T$\wedge$MON) = cr(T$\wedge$TUE) = $\frac{1}{4}$. And we saw in section 7.5 that if Beauty is then told it is Monday, her credences become cr(H) = $\frac{2}{3}$ and cr(T) = $\frac{1}{3}$.

What should her credences be when she wakes up on Tuesday in the tails world? Let's look at her prior *des*, right before she goes to sleep on Monday night. She's been told that it's Monday, so she knows that one of two possibilities obtains: H∧MON or T∧MON. If T∧MON obtains, then her temporal successor will wake up on Tuesday morning with no memory of having woken before: call this evidence e. And if H∧MON obtains, then her temporal successor will wake up on Tuesday morning, with the memory of having woken up before: call this evidence f.

These two successors will be in her *des*. Her *des* also includes the centered worlds she occupies when she wakes up on Monday, since they get evidence e as well. So her *des* will be:

~~TUE(f)~~	TUE(e)
MON(e)	MON(e)
H	T

Her evidence e eliminates one of the heads possibilities, so her credence in heads and tails becomes $\frac{1}{2}/\frac{1}{2}$. So in the case where Beauty is told what day it is shortly after waking, ESC provides Beauty with perfect guidance when she wakes up—she knows she should have an equal credence in heads and tails.

What is Beauty's *des* right before she's told what day it is on Tuesday in the tails worlds? At this point she thinks that one of three possibilities obtains: H∧MON, T∧MON or T∧TUE. If either of the Monday possibilities obtain, then her successor will be told it's Monday. If the Tuesday possibility obtains, then her successor will be told it's Tuesday. So her *des* is:

MON("MON")	TUE("TUE")
	MON("MON")
H	T

Given this *des*, if *per impossibile* she were to learn it was Monday, she would cross out one of the tails possibilities, and her credences would become $\frac{2}{3}/\frac{1}{3}$ according to ESC. So in this case ESC also provides Beauty with perfect guidance when she's told what day it is: she knows that if she learns it's Monday, her credences should become cr(H) $= \frac{2}{3}$ and cr(T) $= \frac{1}{3}$.[8]

But we should not let these local victories distract us from the real moral. The original sleeping beauty is a case of memory loss—Beauty's memories of her prior credences are erased. And any rule which has the form of classical Bayesianism—which employs one's previous credences and evidence to generate one's new credences—will usually be of little use as a source of guidance in these kinds of cases. To some extent we can mitigate this to some extent by employing rules which use less of the information encoded in one's prior credences, but this is a stop-gap measure at best. After all, if the subject can't remember anything about her prior credences, then no rule of this kind can provide her with guidance.

7.6.4 Comparing ESC and CeC$_L$

ESC and CeC$_L$ are similar in a number of respects. They assign similar credences to sleeping beauty when she wakes up on Monday, they offer the same responses to the three skeptical scenarios, and so on. Given this similarity in their assignments, it's natural to ask how these two accounts compare.

CeC$_L$ seems to have an advantage over ESC with respect to guidance. Unlike ESC, CeC$_L$ offers perfect guidance to Beauty in the standard sleeping beauty case. And CeC$_L$ offers guidance to subjects who have forgotten recent facts about the world, like the outcomes of coin tosses. In chapter 8 we'll see that things are a bit more complicated. But for now, we can just note that CeC$_L$ seems to be more useful as a guidance tool than ESC.

ESC is well motivated: it's egalitarian and satisfies LP$_1$. And unlike CeC$_L$,

[8] Of course, since her evidence would also inform her of what her prior *des* really was, ESC would provide perfect guidance in this case no matter what. I.e., even if her pre-telling Tuesday *des* was different from her pre-telling Monday one, and so yielded different credences, it wouldn't lead to a guidance failure, since Beauty's evidence—that it's Monday—would entail that her Tuesday *des* isn't the right one to use.

ESC doesn't need to employ seemingly arbitrary prior constraints like the No Increase Principle.

How well does each fare with respect to the Learning Principles? Both treat the narcissistic scientists case in the same way. As a result, both violate LP_n.[9] We've seen that ESC does satisfy LP_1, however (see Appendix 9.19). What about CeC_L?

To assess CeC_L in this regard, it will be helpful to consider some cases involving fusion. The natural way to think about fusion is as the inverse of fission: instead of one individual splitting into two (or more), two (or more) individuals merge into one. But fusion raises some tricky issues that we've avoided in our discussions of fission. Consider two individuals, with credences cr_1 and cr_2, respectively, who are fused together. What credences should this "fusile" have according to a rule like classical Bayesianism, TSC or ESC? These rules generate the subject's new credences from her evidence and her previous credences. But it's not clear how to apply these rules in this case, since the fusile has two distinct prior credences.

Similar issues arise with respect to fusion for hypothetical prior rules. If two individuals with different hypothetical priors are fused together, what credences should CeC_E, CeC_L and CoC_M assign the fusile? Should these rules pick one or the other, use some mixture of the two, or simply remain silent about such cases?

Fortunately, given our concerns, we don't need to resolve these matters here. Just as we've restricted our attention in cases of fission to those where the fissiles have the same credences and/or hypothetical priors, we will restrict our attention in cases of fusion to those where the individuals have the same credences and/or hypothetical priors. (The claims I make about TSC and ESC in this chapter, and the proofs regarding them given in the appendix, are implicitly restricted to cases

[9] Though, as we'll see in section 7.7, on the natural way of understanding ESC, ESC and CeC_L will diverge in their treatments of the sadistic scientists case.

that do not involve the fusions of individuals with different credences.)

Recall the fission version of the sleeping beauty case: if the coin toss landed tails then Beauty will be fissioned on Sunday night, and both of her fissiles will be put in subjectively indistinguishable situations. Now add the following twist: a minute after the two fissiles wake up, they will be fused back into single person, all while being kept in the same subjective state as Beauty would be in had the coin landed heads.

What should Beauty's credences be according to TSC and ESC? (Since Beauty's *dts* is the same as her *des*, TSC and ESC treat this case the same way.) As we've seen, when she wakes up her credence in heads and tails will be $\frac{1}{2}/\frac{1}{2}$. Right before a minutes has passed, her *dts*/*des* will consist of a single temporal successor at each world:

Original(e)	Fusile(e)
H	T

Since both possibilities are compatible with her evidence e, her credence in heads and tails remains $\frac{1}{2}/\frac{1}{2}$.

What will her credences in this case be according to CeC$_L$? The argument given in section 5.6 with respect to the original sleeping beauty case applies in an identical fashion here: when Beauty wakes up her credences are cr(H) = cr(T) = $\frac{1}{2}$, and cr(T∧MON(9)∧Fiss1) = cr(T∧MON(9)∧Fiss2) = $\frac{1}{4}$.

What should her credences be a minute later? CeC$_L$'s Continuity Principle entails that:

$$\frac{hp(\mathrm{T} \wedge \mathrm{MON}(9) \wedge \mathrm{Fiss1})}{hp(\mathrm{H} \wedge \mathrm{MON}(9))} = \frac{hp(\mathrm{T} \wedge \mathrm{MON}(9{:}01) \wedge \mathrm{Fusile})}{hp(\mathrm{H} \wedge \mathrm{MON}(9{:}01)}. \tag{7.6}$$

Since on CeC, a subject's credences are proportional to her priors, it follows that:

$$\frac{hp(\mathrm{T} \wedge \mathrm{MON}(9) \wedge \mathrm{Fiss1})}{hp(\mathrm{H} \wedge \mathrm{MON}(9))} = \frac{1}{2}, \tag{7.7}$$

and thus that:

$$\begin{aligned} \frac{hp(\text{T} \wedge \text{MON(9:01)} \wedge \text{Fusile})}{hp(\text{H} \wedge \text{MON(9:01)}} &= \frac{cr(\text{T} \wedge \text{MON(9:01)} \wedge \text{Fusile})}{cr(\text{H} \wedge \text{MON(9:01)}} \\ &= \frac{1}{2}, \end{aligned} \tag{7.8}$$

so Beauty's credences become:

$$cr(\text{T} \wedge \text{MON(9:01)} \wedge \text{Fusile}) \;=\; \frac{1}{3}, \;\; \text{and} \;\; cr(\text{H} \wedge \text{MON(9:01)} = \frac{2}{3}. \tag{7.9}$$

On CeC$_L$ Beauty's credence in heads and tails will inevitably change from $\frac{1}{2}/\frac{1}{2}$ right before a minute has passed to $\frac{2}{3}/\frac{1}{3}$ right after. So she will violate LP$_1$. And since LP$_1$ is a special case of LP$_n$, she will violate LP$_n$ as well.

As with the other Learning Principle violations, a little work turns this into a skeptical scenario. We need only (i) make the number of fissiles created and fused large enough, and/or make the fission/fusion process a reoccurring one, (ii) put Beauty in a situation where she should have a very high credence that such a process will take place, and then (iii) show that CeC$_L$ requires Beauty to become virtually certain that such a process is not taking place.

So although CeC$_L$ and ESC are similar, CeC$_L$ is not as successful at accommodating the Learning Principles. Like CeC$_E$ and CoC$_M$, CeC$_L$ violates both of the Learning Principles.

7.6.5 Comparing ESC and TSC

How does TSC compare to ESC?

TSC has the advantage of satisfying both Learning Principles. ESC satisfies LP$_1$, but does not satisfy LP$_n$, and this lapse leaves it susceptible to the narcissistic scientists argument. On the other hand, ESC provides an egalitarian treatment of different cases of self-location. As we've seen, it assigns the same credences to Beauty on Monday in the duplication, fission and original versions of the sleeping beauty case. And ESC is somewhat better as a guidance tool than TSC.

As far as which account one should prefer, I think one's feelings about the narcissistic scientists case provide a decent litmus test. First, one might feel that ESC's treatment of the standard narcissistic scientists case is counterintuitive, but TSC's treatment of the fission version of the narcissistic scientists case is not. On this way of seeing things, TSC's inegalitarianism is not a demerit, it's a mark in its favor. Intuitively, cases of fission *are* different from cases of duplication and cases of temporal location. And a satisfactory account of rational belief change should capture this distinction. These sentiments favor TSC over ESC.

Second, one might feel that neither the original narcissistic scientists case nor its fission variant is counterintuitive. It's the other two skeptical scenarios—the many brains and varied brains cases—that are intuitively unacceptable. On this way of seeing things, TSC's inegalitarianism is likely to seem implausible. And the fact that TSC accommodates LP_n and avoids the narcissistic scientists argument is not really a mark in its favor. These sentiments favor ESC over TSC.

Finally, one might feel that both versions of the narcissistic scientists argument are equally counterintuitive. If one feels this way, then one's animosity towards the narcissistic scientists case cannot just be due to its violation of LP_n—some other intuition is coming into play. I think the likely culprit is something we already saw in chapter 5—the persistent intuition that evidence cannot justify changing our beliefs about the world unless it adds or eliminates some of our doxastic worlds. These sentiments probably favor ESC over TSC, but someone with these inclinations is unlikely to be happy with either. In that case, CoC or CoC_M will probably be most appealing.

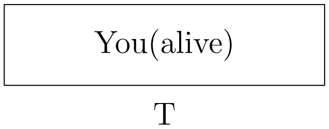
7.7 Death

7.7.1 Two Ways of Treating Death

Interesting issues arise when we consider how successor conditionalization accounts like TSC and ESC ought to treat death. There are two natural ways for such an account to treat death. The way I've set things up delivers one of these treatments—I'll call this the "canonical" treatment—but as we'll see in section 7.7.2, TSC and ESC can be modified to yield an "alternative" treatment as well.

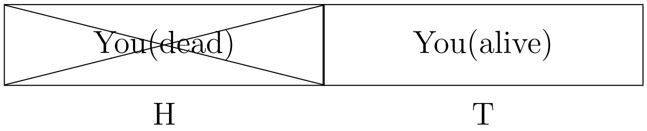
Since these issues are largely orthogonal to the differences between TSC and ESC, I'll avoid the hassle of looking at the two accounts separately by only considering cases which they treat in the same way. Let's start with a simple case of "world-death"—a case where all of your successors at some doxastic worlds die. Consider a case where your life depends on the outcome of a coin toss: if the coin lands heads, you'll be killed instantly; if the coin lands tails, you'll survive. How should the SC accounts treat this case?

Here is the canonical way to proceed. Your *des* right before the coin toss is this:



The heads worlds don't appear in your *des* because you won't have any temporal successors at the heads worlds: you'll be dead. If you find yourself alive, and thus in a position to have credences, your evidence won't eliminate any of these possibilities, and your credence in tails will be 1.

Here's another way way to proceed. We might try to represent the space of your doxastic successors right before the coin toss as:



If you wake up to find yourself alive, then you'll eliminate the heads possibility, and so your credence in tails will become 1.

In standard cases of world-death like this one, both approaches lead to the same credence assignments. But this is not the case with "partial world death"—cases in which only some of your successors at your doxastic worlds die. Consider a variant of the fission version of the sleeping beauty case, where a minute after Beauty wakes up, the second fissile (if such there be) will be disintegrated. On the canonical way of treating this case, Beauty's doxastic successors right before a minute has passed will be:

Original(alive)	Fissile 1(alive)
H	T

If she finds herself alive, her evidence won't eliminate any of these possibilities, and her credences will remain $cr(\text{H}) = cr(\text{T}) = \frac{1}{2}$.

On the alternative way of treating death suggested above, her doxastic successors right before a minute has passed will be:

Original(alive)	~~Fissile 2(dead)~~
	Fissile 1(alive)
H	T

If she finds herself alive she'll eliminate the possibility of being the second fissile, and come to have credences of $cr(\text{H}) = \frac{2}{3}$ and $cr(\text{T}) = \frac{1}{3}$.

So in cases of partial world death, the difference between these two ways of treating death becomes relevant. We'll assess the relative merits of these different treatments in a moment. But let's first look at how one might tweak the SC accounts if one prefers the alternative treatment of death.

7.7.2 How to Implement the Alternative

How might we tweak the SC accounts in order to obtain the alternative treatment?

The problem, from this perspective, is that your dead successors aren't included in your space of doxastic successors. And this stems from the characterizations laid out in chapter 4. In particular, dead people are barred from being temporal successors in two ways.

1. Temporal continuants (and thus successors) are partially characterized in terms of the personal identity relation, which is required to hold between epistemic subjects (i.e., belief-having centered worlds).

2. A subject's temporal successors are essentially her temporal continuants that are in her next subjective state.[10] But subjective states are only had by epistemic subjects.

(This also bars dead people from being *epistemic* successors: since dead people can't have subjective states, they can't ever be in the same subjective state as one of your temporal successors.)

One way to remedy this is to change some of these characterizations to allow for dead successors. For example, to address 1 we might change the relation between temporal continuity and personal identity from a biconditional into a conditional—$PI(a, b) \Rightarrow C_t(a, b) \vee C_t(b, a)$—allowing for dead continuants. And to address 2, we might allow dead people to have a "null" subjective state.

But these kinds of changes lead to headaches. Our notion of temporal continuity was already a bit fuzzy, and the notion gets substantially slippier once we allow for dead continuants. And an even bigger headache arises when we try to determine one's epistemic successors. If all dead people have "null" subjective states, then our dead successors will be in the same subjective state as everyone else's dead successors. But this means that whenever a successor dies at one of our

[10] I say "essentially" because strictly speaking this isn't right. If your next three subjective states are a, then b and then a again, only the temporal continuants who are in subjective state a the first time will be temporal successors.

doxastic worlds, our *des* will be flooded with centered worlds focused on the dead. Likewise, if dead people have "null" subjective states, it's hard to see why rocks and other inanimate objects wouldn't have them as well. If so, then whenever a doxastic successor of ours dies, every inanimate object in each of our doxastic worlds will become an epistemic successor, and any semblance of well-behaved credence evolution will vanish.

A better option is to leave our earlier characterizations the same, and instead amend TSC and ESC directly. In the case of TSC, we can replace *dts* with "*dts**": your *dts* plus a surrogate successor for each of your doxastic alternatives that doesn't have a temporal successor. Likewise, in the case of ESC we can replace *des* with "*des**": your *des* plus a surrogate successor for each of your doxastic alternatives that doesn't have a temporal successor. Both of these amendments will yield the alternative treatment of death described above.

7.7.3 Assessing the Alternatives

We've seen that we can modify TSC and ESC to yield the alternative treatment of death. But should we want to amend the SC accounts in these ways? What reasons are there for adopting one of these treatments of death over the other?

Let's start by looking at two symmetry arguments for the alternative treatment of death. Here is one argument. Consider two cases involving a fair coin toss. In the first case, nothing happens if the coin lands heads. If the coin lands tails, there's a 0.99 chance that you will be disintegrated. In the second case, nothing happens if the coin lands heads, as before. But if the coin lands tails, you will be fissioned into 100 fissiles, and all but one of them will be disintegrated. Now, suppose you find that you're alive. What should your credences be in the outcome of the coin toss?

The alternative stance treats both cases in exactly the same way, while the canonical stance does not. In the first case, both assign you a credence of $cr(\mathrm{H})$

$= \frac{100}{101}$ and $cr(\mathrm{T}) = \frac{1}{101}$. What about the second case? On the alternative way of treating death, when you find yourself alive you will eliminate a number of tails possibilities, and so your credence in heads will increase in the same way as in the first case: $cr(\mathrm{H}) = \frac{100}{101}$ and $cr(\mathrm{T}) = \frac{1}{101}$. But on the canonical treatment of death, finding yourself alive doesn't eliminate any of your previous doxastic successors, so your credences in the second case will remain $cr(\mathrm{H}) = cr(\mathrm{T}) = \frac{1}{2}$.

According to the proponent of the alternative treatment, cases of world-death aren't different from cases of successor-death. So our credence in the outcome of the coin toss in the two cases should be the same. Since the canonical treatment does not yield this result, it should be rejected.

Here is the second argument. Again, consider two cases involving a fair coin toss. In the first case, you will be put in a black room if the coin lands heads. If the coin lands tails, then you will be fissioned into a hundred fissiles; one will be put in a black room, and the other 99 will be put in white rooms. In the second case, you will be put in a black room if the coin lands heads, as before. If the coin lands tails, then you will be fissioned into a hundred fissiles; one will be put in a black room, and the other 99 will be disintegrated. Now, suppose you find yourself in a black room. What should your credences be in the outcome of the coin toss?

Again, the alternative stance treats both cases in exactly the same way, while the canonical stance does not. In the first case, both assign you a credence of $cr(\mathrm{H}) = \frac{100}{101}$ and $cr(\mathrm{T}) = \frac{1}{101}$. What about the second case? On the alternative stance, you will eliminate a number of tails possibilities when you find yourself in a black room instead of dead, so your credence in heads will increase in the same way as in the first case: $cr(\mathrm{H}) = \frac{100}{101}$ and $cr(\mathrm{T}) = \frac{1}{101}$. But on the canonical stance, finding yourself in a black room won't eliminate any of your previous doxastic successors, so your credence in heads and tails will remain $cr(\mathrm{H}) = cr(\mathrm{T}) = \frac{1}{2}$.

According to the proponent of the alternative treatment, we should treat the evidence that you're alive (not dead) in the same way as the evidence that you're in a black room instead of a white one. Of course, both ways of handling death do treat these kinds of evidence in the same way in cases of world-death. But shouldn't we treat these kinds of evidence in the same way in cases of successor-death as well? If so, our credence in the outcome of the coin toss in the two cases should be the same, *pace* the canonical treatment.

What reasons might one give in support of the canonical treatment of death? The reasons given for the alternative treatment claimed that each of the pair of cases described above should be treated in the same way. But a number of people have held that the opposite is true: in each argument the pair of cases are, intuitively, very different.

The bulk of this discussion has taken place in the context of the Many-Worlds interpretation of quantum mechanics—where quantum mechanical chance events are essentially replaced with universe-level fission. I think there's little to be found at the level of analytic argument here on either side. Like the two symmetry arguments given above, the arguments offered in favor of something like the canonical treatment tend to be plausibility arguments, not deductive ones. But some quotes may help give us a feel for the sentiments of those who are likely to prefer the canonical treatment.

The first is by David Lewis:

> "What should [Schrodinger's cat] expect to experience, if it's a very smart cat and knows the set-up, and if it knows there are no collapses? The intensity rule says: expect branches according to their intensities. The intensities are equal. So the cat should equally expect to experience life and death.
>
> But that's nonesense! There's nothing it's like to be dead. Death is oblivion. (Real death I mean. Afterlife is life, not death.) The experience of being dead should never be expected to any degree at all, because there is no such experience. So it seems the intensity rule does not work for the life-and-death branching that the cat undergoes. ...

> When we have life-and-death branching, the intensity rule as so far stated does not apply. We must correct it: first discard all the death branches, because there are no minds and no experiences associated with death branches. Only then divide expectations of experience between the remaining branches in proportion to their intensities. ... The cat should expect with certainty to find itself still alive after the evil experiment, since that is the guidance delivered by the corrected intensity rule."[11]

The second comes from Huw Price:

> "...these consequences expose the no collapse view to an unusual form of empirical verification. For suppose we play Russian roulette, quantum style, and find ourselves surviving long after the half-life predicted by orthodox views. This would be very good evidence that the no collapse view was correct. If we wanted to share our evidence with skeptical colleagues, we would need to ensure that they too participated in the game of course—otherwise they would be left saying "I told you so" to our corpse in most of their surviving branches. (Alternatively, we might persuade each of our colleagues to participate in his or her own private game, and point out that the view predicts that each will find that he or she does very much better than average—indeed, that he or she survives everybody else in the initial group!)"[12]

So which way of treating death should we adopt? The canonical treatment is more natural given the way TSC and ESC are constructed. And my intuitions tend to side with the canonical approach. But having talked to many who feel the opposite way, I'm diffident about this.

Although the question of how to treat death is an interesting topic, it's largely orthogonal to the other issues of interest. So for the purposes of this work, I won't take a stand of the matter. Instead, I'll leave these open as other versions of successor conditionalization one could adopt. So we end this chapter with these four successor conditionalization accounts to work with, each employing a different space of doxastic successors.

[11] From a pre-print of Lewis (2004), p.14-15.

[12] From Price (1996), p.222.

Chapter 8
Guidance

8.1 Introduction

Our first attempt to extend classical Bayesianism to accommodate self-locating beliefs resulted in the *hp*-accounts of chapter 5. But as we saw in chapter 6, these accounts fail to satisfy some powerful intuitions about when belief change is justified: the Learning Principles. In chapter 7 we constructed accounts to satisfy these principles: TSC and ESC. But we saw that in some cases they assign different credences to subjects in the same subjective state. As a result, TSC and ESC have limited use as guidance tools, since a subject won't always be in a position to determine the credences TSC or ESC assigns.

We'd like an account that both provides guidance and satisfies the Learning Principles. But there is a deep tension between guidance and the Learning Principles. And this tension requires us to make some uncomfortable choices.

We looked at the extent to which TSC and ESC provide guidance in the last chapter. In the next section we'll look at the extent to which the *hp*-accounts provide guidance. In the third section we'll look at what a subject should do when her updating rule doesn't provide her with guidance. In the fourth section we will look at the source of the tension between guidance and the Learning Principles. We'll conclude in the fifth section by assessing the options available to us in light of this conflict.

8.2 Guidance and the Hypothetical Prior Accounts

We can think of an updating rule R as a function that takes arguments and spits out the credences a subject ought to have. Classical Bayesianism, TSC and ESC take a subject's evidence and her previous credences as arguments.[1] *Hp*-conditionalization and the *hp*-accounts of chapter 5 take a subject's evidence and her hypothetical priors as arguments.

An account of belief updating fails to provide guidance in situations where what the subject has access to isn't sufficient to determine the credences that the account assigns her. (By "has access to" I mean what the subject can, in theory, deduce given only her subjective state. A subject always has access to her evidence, for example.) Classical Bayesianism fails to provide guidance in the case where a subject forgets the outcome of a coin toss because she doesn't have access to the crucial argument: her previous credence in heads and tails. TSC fails to provide guidance in the sleeping beauty case because Beauty, upon waking on Monday, doesn't know enough about the arguments TSC employs—in particular, her prior credences—to determine the credences it assigns her.

What about the *hp*-accounts of chapter 5? The extent to which they provide guidance depends on how we understand hypothetical priors. In chapter 1 we saw three ways to make sense of hypothetical priors:

(1) the initial credence of a subject before she's received any evidence,

(2) the "normative stamp" on a subject indicating what credences she ought to have if she had no evidence,

(3) any probability function that would make a subject's credences at different times satisfy the normative account in question.

[1] Don't we need to add arguments encoding her previous *dts* and *des*? No: as we saw in chapter 4, these are determined by her prior credences.

Suppose we understand hypothetical priors along the lines of (2)—as normative stamps on subjects. If we're objective Bayesians, and so we think everyone has the same normative stamp, then it's natural hold that we have *a priori* access to our hypothetical priors. I.e., we might think that sufficient thought will reveal Indifference Principles which will uniquely determine what the credences of a subject who is maximally indifferent—a subject who knows nothing—should be. On this understanding a subject always has access to her hypothetical priors, and the *hp*-accounts always provide guidance.

Suppose, though, that we take these normative stamps to be something we don't have guaranteed access to—perhaps they supervene on our inductive dispositions, for example, but we can only figure out what our inductive dispositions are by looking at how our beliefs have changed in the past. Or suppose we understand hypothetical priors along the lines of (1) or (3), in which case we also aren't guaranteed access to them. On these veiled ways of understanding hypothetical priors, the *hp*-accounts often provide more guidance than the SC accounts.

Let's look at some examples. First, consider the memory loss case described above, where a subject observes the outcome of a coin toss and then forgets this outcome. The subject doesn't have access to her prior credence in the outcome of the coin toss, but she does have access to her credences at earlier times, so she still has access to her hypothetical priors. In this case the *hp*-accounts provide the subject with guidance, while the SC accounts do not.

Likewise, the *hp*-accounts provide Beauty with guidance when she wakes up in the original sleeping beauty case. Although she is no longer certain of her prior credences, she still knows what her credences were before Sunday night, and so still has access to her hypothetical priors.

But the *hp*-accounts often fail to provide guidance in many of the same kinds of situations as the SC accounts. Consider a variant of the memory loss case where the subject forgets not only her previous credence in heads, but her credence in heads at all earlier times as well. In this case the subject won't have access to her veiled priors. And so, like the SC accounts, the *hp*-accounts won't provide her with guidance.

Likewise, consider the following variant of the sleeping beauty case. In this version, instead of being woken up twice if the coin lands tails, a stranger will be brainwashed and put in a subjective state identical to Beauty's Monday morning state. Suppose the stranger's hypothetical priors are different from Beauty's. When Beauty wakes up on Monday morning, she doesn't know whether she is Beauty or the stranger, and thus she doesn't know what her hypothetical priors are. So, like the SC accounts, the *hp*-accounts won't provide Beauty with guidance.

Next consider the duplication variant of the sleeping beauty case. The duplicate is newly created, and so won't have any prior credences. On the (1) way of understanding hypothetical priors, the duplicate doesn't have priors yet. On the (3) way of understanding hypothetical priors, the duplicate's "hypothetical priors" can be any probability function, since her prior credences don't constrain it. So given (1) or (3), the *hp*-accounts won't provide guidance in cases of duplication, since they'll impose different prior constraints on Beauty and the duplicate.

So the veiled *hp*-accounts, like the SC accounts, fail to provide guidance in cases of memory loss, in sleeping beauty-like cases, in cases of duplication, and so on.

Finally, there are some cases where the SC accounts provide guidance and the veiled *hp*-accounts do not. Consider a case where you remember your previous credences, but you don't remember your credences at any other time. Furthermore, you don't know how you arrived at your previous credences: you might have gotten them by updating in the appropriate manner (using your hypothetical priors and evidence, say), or you might have been induced by a foreign power to adopt credences that are radically different from the ones you ought to have adopted. In this case the SC accounts will provide guidance, since you know what your previous credences are. But the veiled *hp*-accounts will not provide guidance, since you have no way of deducing what your hypothetical priors are.

When all is said and done, though, the *hp*-accounts do better than the SC accounts with respect to guidance. This is because it's easier to get indirect access to your hypothetical priors than your previous credences. If you know your previous credences and your evidence, for example, you can usually deduce your hypothetical priors. So the *hp*-accounts generally provide guidance in cases where the SC accounts provide guidance. And since the *hp*-accounts also provide guidance in cases where you have access to your hypothetical priors but not your previous credences, the *hp*-accounts will generally provide more guidance than the SC accounts.

If we take hypothetical priors to be something a subject always has access to then this advantage is dramatic: the *hp*-accounts always provide guidance. If we adopt a veiled version of the *hp*-accounts, the advantage is much smaller, though still significant. But these accounts will fail to provide guidance in many of the same kinds of cases as the SC accounts, and in some cases they won't provide guidance even though the SC accounts will. So while the veiled *hp*-accounts have an advantage over the SC accounts with respect to guidance, it is not an overwhelming one.

8.3 Rules of Thumb for Guidance

Suppose you're a proponent of ESC, and you're put in the sleeping beauty case. As we saw in section 7.6.3, ESC assigns you a credence of $\frac{1}{2}$ in heads on Monday morning, and a credence of $\frac{3}{7}$ in heads on Tuesday morning if the coin flip lands tails. Since you don't know whether it's Monday or Tuesday when you wake up, you don't know which credence to adopt.

But even though ESC fails to provide you with guidance, you still might think that certain credences are more reasonable than others, in light of ESC's assignments. For example, it might seem unreasonable to adopt $cr(\text{H}) = 1$, given that you know ESC assigns you either $cr(\text{H}) = \frac{1}{2}$ or $cr(\text{H}) = \frac{3}{7}$. Rather, it seems natural to adopt a credence in the interval $[\frac{3}{7}, \frac{1}{2}]$. So even in cases where a rule itself fails to provide guidance, it seems there might be additional "rules of thumb" regarding guidance.

This possibility is intriguing. The main reason for our dissatisfaction with the *hp*-accounts of chapter 5 is that they fail to satisfy either of the Learning Principles. The SC accounts satisfy at least one of the Learning Principles, but fail to provide guidance in a number of key cases. But if we can get some guidance in these cases by employing rules of thumb, then perhaps we can get the best of both worlds: a set-up which both provides us with guidance and satisfies some of the Learning Principles.

This possibility also raises some questions. If there are rules of thumb regarding guidance, then it seems we have two layers of normative advice: the advice of the updating rules and the advice given by the rules of thumb. But why are there two layers? What distinguishes them?

How we answer these questions depends on what we take an account of belief dynamics to be. One stance is this: the purpose of an account of the dynamics of belief is to give a subject a means to determine the credences she should adopt. On this view, there is no room for two layers of normative advice. If adding these rules of thumb leads to an improvement over the original account, then this is an indication that the account was flawed, and that the rules of thumb should be incorporated into the account directly. So the proposal to adopt rules of thumb regarding guidance is a proposal to modify the account in question.

Another stance is that an account of belief dynamics is nothing more than a diachronic credence constraint: it tells us how a subject's credences at different times ought to be related. Now, if a subject has access to her prior credences (actual or normative[2]), then she can use the account to figure out how her current credences should cohere with them. But this account won't always provide subjects with guidance. It's not reasonable, after all, to expect an account to tell a subject how to line up her current beliefs with her prior beliefs if she doesn't have access to her prior beliefs.

On this view there is room for two layers. The purpose of the account is to spell out how credences at different times should be related if they are to cohere with each other in an ideal way. And although we can use the account for guidance in some cases, this isn't its primary purpose. This is the purpose of the rules of thumb: they give subjects advice about what credences to adopt in cases where the account does not.

On either stance, the possibility of guidance rules of thumb is of interest. So what kinds of rules of thumb might one adopt?

It depends on which account we start with. Any account which suffers from guidance failures can be used in conjunction with rules of thumb regarding guidance. The SC accounts are particularly interesting, because they satisfy at least one of the Learning Principles. By adding guidance rules of thumb to the SC accounts, there's the intriguing promise of getting both LP-satisfaction and guidance. In contrast, the veiled *hp*-accounts hold less promise, since they don't satisfy the Learning Principles.

Let's look at how one might add some rules of thumb to an SC account. I'll

[2] Actual in the case of accounts like classical Bayesianism, normative in the case of accounts which employ hypothetical priors.

use ESC as my example, but one could add similar rules to TSC just as well. A natural starting point is this:

RoT$_1$ (permissible): If possible, adopt credences that you know are permitted by the account.

This rule of thumb aids ESC in the duplication version of the sleeping beauty case. Recall ESC's treatment of this case, described in section 7.6.3. ESC assigns Beauty a credence of $\frac{1}{2}$ in heads, but fails to constrain the credences of the duplicate since the newly created duplicate has no prior credences. And since Beauty doesn't know whether she's the duplicate, ESC doesn't provide her with guidance. But RoT$_1$ does provide guidance: it recommends that Beauty adopt $cr(\mathrm{H}) = \frac{1}{2}$, since ESC permits this whether or not she's the duplicate.

However, RoT$_1$ won't help in the case we brought up at the beginning of this chapter, the original sleeping beauty case. In that case Beauty knows that ESC permits either $cr(\mathrm{H}) = \frac{1}{2}$ or $cr(\mathrm{H}) = \frac{3}{7}$, but she doesn't know which. A natural suggestion in this case is:

RoT$_2$ (span): If there aren't any credences that you know are permitted by the account, then adopt a credence in the span of the values that might be permitted.

Since ESC assigns Beauty either $cr(\mathrm{H}) = \frac{1}{2}$ or $cr(\mathrm{H}) = \frac{3}{7}$ in the original sleeping beauty case, RoT$_2$ recommends that she adopt some credence in heads such that $cr(H) \in [\frac{3}{7}, \frac{1}{2}]$.

ESC together with these two rules of thumb provide guidance in every case (though this won't always be *precise* guidance). But the combination of ESC, RoT$_1$ and RoT$_2$ violates LP$_1$. Recall from section 6.2 how LP$_1$ applies to permissive accounts: for a permissive account to satisfy LP$_1$, the subject's prior credence must lie in *every* span of credences that the account permits her to adopt given her doxastic evidence. As we've just seen, when Beauty gets the waking evidence e, this combination allows her to adopt any credence in heads that lies in the interval $[\frac{3}{7}, \frac{1}{2}]$. Since this allows Beauty to adopt a credence in heads that doesn't equal her prior credence—she can adopt $cr_e(H) = \frac{3}{7}$, even though her prior credence is $cr(H) = \frac{1}{2}$—this combination violates LP$_1$.

To address this worry, we might adopt a third rule of thumb:

RoT$_3$ (LP$_1$-satisfaction): If RoT$_2$ permits a choice between a number of credences, choose one which would satisfy LP$_1$, when possible.[3]

How does the addition of RoT$_3$ bear on Beauty's credences? When Beauty wakes up and gets e, RoT$_3$ recommends that she adopt some $cr_e(H) \in [\frac{3}{7}, \frac{1}{2}]$ that would satisfy LP$_1$, if any. Is there such a credence? And if so, what is it?

We've already seen that on Monday morning the combination must assign $cr_e(H) = \frac{1}{2}$ in order to satisfy LP$_1$ with respect to her Sunday night credences. Now let's see whether assigning Beauty a credence of $cr_e(H) = \frac{1}{2}$ on Tuesday morning satisfies LP$_1$ with respect to her Monday night credences.

We're assuming that $cr_e(H) = \frac{1}{2}$, so Beauty's credence in heads when she gets evidence e on Monday morning is $\frac{1}{2}$. Since she doesn't get any relevant evidence between Monday morning and Monday evening, her credence in heads on Monday evening will be $\frac{1}{2}$ as well.

And it's trivially the case that Beauty's Monday night credence in heads will lie in the span of the credences she might be assigned on Tuesday morning if she's assigned $cr_e(H) = \frac{1}{2}$ on Tuesday morning.[4] So this assignment will satisfy

[3] That is, choose the credence that would, if adopted as the assignment of a combination account including RoT$_3$, lead to LP$_1$ being satisfied by the account in the case in question.

[4] Where we're restricting ourselves to possibilities the subject has a positive credence in.

There's a slight complication in this case because on Monday night she has two kinds of doxastic evidence, e and f (where f is the evidence she gets when she wakes up and does remember waking up before). (Recall the diagram of her *des* from section 7.6.3.) But it doesn't matter: since $cr_e(H) = \frac{1}{2}$, $cr_f(H)$ can be anything (it happens to be $\frac{3}{5}$), and it will still be the

LP$_1$ with respect to her credences on Monday night as well. Since $\frac{1}{2}$ is the only credence in heads in $[\frac{3}{7}, \frac{1}{2}]$ that will satisfy LP$_1$, this is what the combination of ESC and RoT$_1$-RoT$_3$ assigns Beauty when she wakes up.

So far it seems this combination gets us what we want: guidance and LP$_1$ satisfaction. But consider the following case:

> *Convergence:* Consider a pair of subjects with the same doxastic worlds, but with different *de dicto* credences. In particular, let the first subject have a credence of $\frac{1}{3}$ in A, and let the second subject have a credence of $\frac{2}{3}$ in A. A minute from now, both subjects will be put into the same subjective state. When they enter into this state, they will have the same *de dicto* information as before—their doxastic worlds will remain the same—but they won't know which subject they are. Both subjects know all this, and know that they have no chance of dying in the near future. What should their credence in A be a minute from now?

According to ESC, each subject should have the same credence in heads as they had before, just as LP$_1$ demands. But this is of little use guidance-wise, since the subjects don't know what their prior credences were. The addition of RoT$_2$ provides the subjects with guidance: it recommends that they both adopt a credence in A in the span of $[\frac{1}{3}, \frac{2}{3}]$. But now LP$_1$ isn't satisfied since this combination allows the subjects to adopt a credence in A that doesn't equal their prior credence. For example, the first subject can adopt $cr_e(A) = \frac{1}{2}$, even though her prior credence is $cr(H) = \frac{1}{3}$. And RoT$_3$ is of little help, since there is no credence in $[\frac{1}{3}, \frac{2}{3}]$ that will satisfy LP$_1$: there is no $cr_e(A)$ equal to both of their prior credences in A. (If the subjects are assigned $cr_e(A) = \frac{1}{3}$, this credence won't equal the second subject's prior credence, $cr(H) = \frac{2}{3}$. If the subjects are assigned $cr_e(A) = \frac{2}{3}$, then this credence won't equal the first subject's prior credence, $cr(H) = \frac{1}{3}$. If the subjects are assigned any other $cr_e(A)$, then this credence won't equal either of their prior credences.)

case that her Monday night credence will lie in the span of the credences she might adopt on Tuesday morning.

So the ESC combination fails to deliver on the intriguing hope that the rules of thumb offered—a combination which would provide us with guidance and ensure that we don't violate LP_1. Is there an account that can?

8.4 Guidance and the Learning Principles

We might hope that the susceptibility of the ESC combination considered above to the convergence case is an artifact of the way we were proceeding. Perhaps if we started with a different account, and considered other kinds of rules of thumb, we would do better. But little about the convergence case depended on details about ESC and the rules of thumb in question. We can set up a generalized convergence argument against an arbitrary account R in the following way:

> *The General Convergence Argument (R):* Consider two subjects who satisfy R, and know the following details of their epistemic situation:
>
> 1. They currently have different *de dicto* credences,
> 2. Their temporal successors will have the same doxastic worlds as they do now,
> 3. Their temporal successors will be in the same subjective state as each other.
>
> If R provides them with guidance, then it must assign the successors the same *de dicto* credences, and so will violate LP_1. If R satisfies LP_1, then it must assign the successors different *de dicto* credences, and so cannot provide guidance. So R cannot both provide the subjects with guidance and satisfy LP_1.

This argument doesn't prove that providing guidance and LP_1 are incompatible—one can deny that the case described is possible—but it does point to a deep tension between the Learning Principles and guidance.

The heart of the tension is this.

In order to satisfy guidance, subjects in the same subjective state must be assigned the same credences. So if we want to ensure that subjects will always have guidance, we want an account's credence assignments to depend only on what they currently have access to—their current subjective state.

The Learning Principles essentially require the credences R assigns you to be appropriately related to your previous credences and the other credences R might have assigned you. If we want to ensure that subjects always satisfy a Learning Principle, we want an account's credence assignments to depend on the subject's previous credences, and the other credences the subject might have adopted. And we want this to be the case regardless of whether the subject has access to this information.

So the two requirements push in different directions: the natural way to satisfy guidance is to adopt a rule whose only argument is one's current subjective state, while the natural way to satisfy the Learning Principles is to adopt a rule whose arguments include one's prior credences and the other credences one might have been assigned.

8.5 Endgame

The ideal account would provide us with guidance and satisfy one or both of the Learning Principles. But, as we've just seen, finding a tenable account which does both will be difficult. So what are our options?

Option 1: Adopt an account which always provides guidance, but sometimes violates the Learning Principles.

One option is to hold firm on our demand that an account always provide guidance, but relax the requirement that it always satisfy the Learning Principles.

If we take this option, we have several choices available from the accounts we've examined. When coupled with the *a priori*-access understanding of hypothetical priors, all of the *hp*-accounts we looked at in chapter 5 provide subjects with guidance.

But the *hp*-accounts lead to skeptical scenarios like the many and varied brains cases. Even though we've relaxed the demand that an account always satisfy LP_1, we might still want to avoid these results. In that case, we might be attracted to the ESC combination account, since it violates the Learning Principles in less radical ways. Similarly, if one would like to satisfy LP_n in most cases, one might look into a combination account starting with TSC.

Option 2: Adopt an account which satisfies some of the Learning Principles, but doesn't always provide us with guidance.

Another option is to hold firm on our demand that an account satisfy LP_1 (and possibly LP_n), but relax the requirement that it provide guidance.

If we take this option, then we have several choices available from the accounts we've examined. We've already seen two accounts that satisfy LP_1: TSC and ESC. And if we want to satisfy both Learning Principles, TSC will satisfy LP_n as well.

Note that if we adopt this option, there are two natural ways to think about guidance. One way is this: guidance failures are inevitable, but they should be minimized. From this perspective, we may or may not be content with TSC or ESC. TSC and ESC do provide a fair amount of guidance—more than classical Bayesianism provides—but we might still want to look around for an account which provides more.

Another way to think about guidance is this: one might reject the idea that guidance is a desiderata of an account of belief dynamics. The notion that guidance is important presupposes doxastic voluntarism: the view that we can choose what beliefs to adopt, just like we can choose what acts to perform. On this picture, an account should provide guidance because it's supposed to help us decide what to believe.

But if we reject doxastic voluntarism, it's natural to reject the importance of guidance as well. On this picture, we don't get to choose what we believe—we just believe it. There's no sense to be made of giving us advice about what to believe if we don't have any control over our beliefs. All we can do is distinguish those people whose beliefs evolve in a coherent, optimal or ideal way, from those whose beliefs do not. And this is what an account of the dynamics of belief does.

Option 3: Try to find an account which provides guidance and satisfies the Learning Principles.

A third option is to persevere, and continue searching for a tenable account which both provides guidance and satisfies one or both of the Learning Principles.

How should we proceed? To start with, we should look for an account that avoids the general convergence argument presented in the last section. Only one kind of account we've looked at escapes this argument: *hp*-accounts that employ CoC (such as CoC_M from chapter 5) together with the objective Bayesian understanding of hypothetical priors discussed in section 8.2 (hypothetical priors as a normative stamp that all subjects share, and that all subjects have *a priori* access to).[5] But this type of account fails to satisfy LP_1, as we saw in chapter 7. So this brings us no closer to our goal.[6]

All said and done, the prospects for this option look bleak.

[5] The proponent of this type of account denies that the general convergence case is possible. In particular, on this account it's impossible for all of the following to hold: (i) the two subjects have different *de dicto* credences, (ii) their successors have the same doxastic worlds as they do, and (iii) their successors are in the same subjective state (and thus have the same evidence). On CoC, a subject's doxastic worlds are the worlds compatible with her evidence that she has a non-zero prior in, and on this account everyone has the same priors, so if (iii)—the successors have the same evidence—then they must have the same doxastic worlds. And if (ii)—the two subjects have the same doxastic worlds as their successors—then the subjects must have the same doxastic worlds as well. But on CoC a subject's doxastic worlds and her hypothetical priors determine her *de dicto* credences, so it follows that the two subjects must have the same *de dicto* credences. And this violates (i).

[6] Is there *any* account that both provides guidance and satisfies the Learning Principles? Yes: consider the account which tells you to always adopt some credence distribution *cr* which you have *a priori* access to, regardless of what your evidence is. This will always provide guidance, and since your credences never change, it will satisfy both Learning Principles. But this account is clearly not tenable.

Chapter 9
Appendix

9.1 Humeanism

Much of the literature on chance has focused on the compatibility of a satisfactory chance-credence principle and Humean supervenience. The theory of chance I propose has little bearing on this issue, as I will show. The majority of this section will look at the impact of adopting an admissibility-free chance-credence principle on the debate over Humeanism. I will end with a quick note on the bearing of the other features of my account on this debate.

Lewis (1994) and others have noted that at worlds where Humean supervenience holds, a chance theory T will generally assign a positive chance to $\neg T$. Consider a simple Humean theory, frequentism. On this account, the chance of a chance event is determined by (i) assigning a chance to outcomes equal to the actual frequency (past and future) of these outcomes, while (ii) treating these events as independent and identically distributed. Now consider a world where frequentism is true, and where there are only two chance events, two coin flips, one which comes up heads and one which comes up tails. Then the chance of a coin flip coming up heads is $\frac{1}{2}$, and the chance of two coin flips coming up heads is $\frac{1}{4}$. But if the coin came up heads twice, then frequentism would assign chance 1 to the coin toss coming up heads. So it seems that Humean chances *undermine* themselves: they assign a positive chance to an outcome on which they wouldn't be the correct chances. More generally, they assign a positive chance to some other chance theory being true.

Given Lewis's Principal Principle, this appears to lead to a contradiction:

$$0 < ch_{TH}(\neg T) = hp(\neg T|\, TH) = 0, \tag{9.1}$$

where the middle equality is furnished by the Principal Principle. On further inspection, this does not lead to a contradiction because Lewis's Principal Principle is equipped with an admissibility clause. The admissibilty clause can be used to disrupt the middle equality of (9.1) and prevent a contradiction. But we only avoid a contradiction by making so much inadmissible that the Principal Principle is useless.

How does the Basic Principle fare? The Basic Principle leads to the same apparent contradiction as the Principal Principle, and since the Basic Principle has no admissibility clause, admissibility cannot be used to disrupt the middle equality of (9.1). So given the Basic Principle, undermining does appear to lead to a contradiction. Regardless of whether we adopt the Principal Principle or the Basic Principle, the Humean seems to be in bad shape.

Lewis (1994) later tried to avoid this problem by adopting an alternate principle:

$$hp(A|\, TH) = ch_{TH}(A|T), \text{ if } hp(A|\, THE) \text{ and } ch_{TH}(A|T) \text{ are defined}(9.2)$$

Since $ch_{TH}(\neg T|T) = 0$ even on a Humean account of chance, adopting (9.2) escapes the problem. A similarly modified version of the Basic Principle avoids the problem in the same way.

In either case, the move Lewis proposes is questionable. Vranas (2002) has shown that the problem that motivated Lewis's adoption of (9.2) is only apparent. Take a world w at which both Humean supervenience S and the chance theory T hold. Let H be an *undermining history* of w relative to T, such that $S \wedge H \Rightarrow \neg T$. T will generally assign a positive chance to w, and so a positive chance to S.

Likewise, T will generally assign positive chances to some histories like H, histories that would entail $\neg T$ if they held at a world where S held. But it doesn't follow from this that T must assign a positive chance to *both* H and S being true, and thus a positive chance to $\neg T$. T can assign a positive chance to H and a positive chance to S while assigning a 0 chance to the conjunction of H and S. So Humean chances don't need to undermine themselves. And this is true regardless of whether the chance-credence principle has an admissibility clause.

(Though how much of a respite this is can be questioned. While the revised theory is compatible with the truth of Humean supervenience at this world, it's incompatible with the more ambitious claim that Humean supervenience is metaphysically or nomologically necessary. And Frank Arntzenius has pointed out that Vranas's treatment still leads to counterintuitive results for subjects who are confident that Humeanism obtains. Given this, it seems none of the Humean responses to the undermining problem are without cost.)

So adopting an admissibility-free chance credence principle has little bearing on the issue of Humeanism. But another feature of my proposal might also seem to be at odds with Humeanism. Namely, the account I offer for the structure of chance theories entails that the measures associated with chance theories are assigned over the worlds where that theory holds. As a consequence, chance theories will always assign themselves a chance of 1. Since it appears that on Humean accounts the chance a chance theory assigns to itself is generally less than one, this seems like an anti-Humean assumption. But as Vranas has shown us, this is a mistake; this assumption is not incompatible with Humeanism. So this feature of my account has little bearing on the issue of Humean supervenience.

9.2 Deriving L5 and L6

Both of these derivations assume the rest of Lewis's metaphysical account (L1-L4) and PP$_2$.

The derivation of L6 also requires the assumption that a chance theory at a world assigns a chance of 1 to the laws that hold at this world. Note that this is more ambitious then the claim that chance theories assign themselves a chance of 1—this requires that *all* of the laws, not just those involving chance, get a chance of 1.

9.2.1 Deriving L5

Let T be a complete theory of chance at a world, and H any history up to a time t at that world. Let E be any proposition about the past (relative to t). Since E is about the past, H entails either E or its negation.

Now, if H entails E, and $ch_{TH}(E)$ is defined, then:

$$\begin{aligned} ch_{TH}(E) &= hp(E|TH) \\ &= \frac{hp(ETH)}{hp(TH)} \\ &= \frac{hp(TH)}{hp(TH)} \\ &= 1 \end{aligned} \tag{9.3}$$

On the other hand, if H entails $\neg E$, and $ch_{TH}(E)$ is defined, then:

$$\begin{aligned} ch_{TH}(E) &= hp(E|TH) \\ &= \frac{hp(ETH)}{hp(TH)} \\ &= \frac{hp(\neg EETH)}{hp(TH)} \\ &= 0 \end{aligned} \tag{9.4}$$

So $ch_{TH}(E) = 1$ or 0. Since this is true for any proposition E about the past, it follows that the past is no longer (non-trivially) chancy.

9.2.2 Deriving L6

Let L be the laws at a deterministic world, and T the complete chance theory at that world. Let H be any history up to a time at this world, and A be any proposition. If L is deterministic, then either $LH \Rightarrow A$ or $LH \Rightarrow \neg A$, since deterministic laws and a complete history up to a time entail everything. Equivalently, either $L \Rightarrow A \vee \neg H$ or $L \Rightarrow \neg A \vee \neg H$.

Suppose $L \Rightarrow A \vee \neg H$. Then:

$$\begin{aligned} 1 &= ch_{TH}(A \vee \neg H) \qquad (9.5) \\ &= hp(A \vee \neg H | TH) \\ &= \frac{hp((A \vee \neg H) \wedge TH)}{hp(TH)} \\ &= \frac{hp(ATH)}{hp(TH)} \\ &= hp(A|TH) \\ &= ch_{TH}(A) \end{aligned}$$

The first line makes use of the assumption that anything the laws entail gets assigned a chance of 1 by the chance laws.

So if $L \Rightarrow A \vee \neg H$, then $ch_{TH}(A) = 1$. If $L \Rightarrow \neg A \vee \neg H$, then an identical derivation yields $ch_{TH}(\neg A) = 1$, i.e., $ch_{TH}(A) = 0$. So for any history H and any proposition A, $ch_{TH}(A) = 1$ or 0. I.e., all of the chance distributions associated with T assign only trivial chances. Since this is true for any chance theory of a deterministic world, it follows that determinism and (non-trivial) chances are incompatible.

9.3 Deriving PP$_1$ from PP$_2$, and Vice Versa

Both of these derivations employ L1-L4 and an assumption about admissibility.

The claim about admissibility Lewis (1986*b*) employs in this derivation is the following:[1] If $\langle ch_t(A) = x\rangle E$ is admissible at t then $\langle ch_t(A) = x\rangle E$ can be expressed as a non-empty disjunction of the grounds of chance distributions, $T_iH_i^t$ such that $ch_{T_iH_i^t}(A) = x$ (where H^t is a complete history up to t). $\langle ch_t(A) = x\rangle$ can always be expressed this way, of course, so this is really a constraint on E: the intersection of E and $\langle ch_t(A) = x\rangle$ must still be equivalent to a disjunction of this kind. If E 'cuts across' these grounds in some way, then the disjunction won't be expressible in this way, and E can't be admissible.

9.3.1 $\mathbf{PP}_2 \Rightarrow \mathbf{PP}_1$

Let $T_1H_1^t$ through $T_nH_n^t$ be all of the different chance theory and complete history up to t pairs such that the chance of A is x. By definition,

$$\langle ch_t(A) = x\rangle = T_1H_1^t \vee ... \vee T_nH_n^t \tag{9.6}$$

If $\langle ch_t(A) = x\rangle E$ is admissible, then:

$$\begin{aligned} hp(A|\langle ch_t(A) = x\rangle E) &= hp(A|E \wedge (T_1H_1^t \vee ... \vee T_nH_n^t)) \\ &= hp(A|T_1H_1^t \vee ... \vee T_mH_m^t) \\ &= \frac{hp(A \wedge (T_1H_1^t \vee ... \vee T_mH_m^t))}{hp(T_1H_1^t \vee ... \vee T_mH_m^t)} \\ &= \frac{\sum_i hp(AT_iH_i^t)}{\sum_j hp(T_jH_j^t)} \\ &= \frac{\sum_i hp(T_iH_i^t) \cdot hp(A|T_iH_i^t)}{\sum_j hp(T_jH_j^t)} \\ &= \frac{\sum_i hp(T_iH_i^t) \cdot ch_{T_iH_i^t}(A)}{\sum_j hp(T_jH_j^t)} \\ &= x \cdot \frac{\sum_i hp(T_iH_i^t)}{\sum_j hp(T_jH_j^t)} \\ &= x. \end{aligned} \tag{9.7}$$

[1] See Lewis (1986*b*), p.99-100.

So $hp(A|\langle ch_t(A) = x\rangle E) = x$ if $\langle ch_t(A) = x\rangle E$ is admissible.

9.3.2 PP$_1$ ⇒ PP$_2$

Suppose $ch_{TH^t}(A) = x$. Then $TH^t \Rightarrow \langle ch_t(A) = x\rangle$, and thus:

$$\begin{aligned} hp(A|TH) &= hp(A|\langle ch_t(A) = x\rangle TH) \\ &= x \\ &= ch_{TH^t}(A), \end{aligned} \tag{9.8}$$

where the next to last step makes use of the admissibility assumption given above.

9.4 PP$_3$ Derivations

As before, these derivations employ L1-L4 and the admissibility assumption given above. We've already seen that, given these assumptions, PP$_1$ can be derived from PP$_2$, and vice versa. So to show that PP$_3$ can also be derived from PP$_1$ and PP$_2$, and vice versa, it will suffice to show that PP$_3$ can be derived from PP$_2$ and PP$_2$ can be derived from PP$_3$.

Recall that the formulation of PP$_3$ makes use of the following definition: $\langle ch(A) = x\rangle E$ is admissible iff (i) $\langle ch(A) = x\rangle E \neq \emptyset$, and (ii) $\forall t$ either (a) $\langle ch_t(A) = x\rangle E$ is admissible relative to t, or (b) $\langle ch_t(A) = x\rangle E = \emptyset$. Given Lewis's assumption about admissibility (see 9.3), this definition entails that $\langle ch(A) = x\rangle E$ will be admissible iff it can be expressed as a disjunction of the grounds of chance distributions, T_iH_i (where the histories do not need to be up to the same time).

9.4.1 PP$_2$ ⇒ PP$_3$

Let t_1-t_n be all of the times t such that, for some complete chance theory T and history H^t, $ch_{TH^t}(A) = x$. Then by definition,

$$\langle ch(A) = x\rangle = \langle ch_{t_1}(A) = x\rangle \vee ... \vee \langle ch_{t_n}(A) = x\rangle. \tag{9.9}$$

If $\langle ch(A)=x\rangle E$ is admissible then:

$$\begin{aligned}
hp(A|\langle ch(A)=x\rangle E) &= hp(A|E \wedge (\langle ch_{t_1}(A)=x\rangle \vee ... \vee \langle ch_{t_n}(A)=x\rangle)) \qquad (9.10)\\
&= hp(A|E \wedge (H_1T_1 \vee ... \vee H_mT_m))\\
&= hp(A|(H_1T_1 \vee ... \vee H_qT_q))\\
&= hp(A|(H_1T_1 \vee ... \vee H_rT_r))\\
&= \frac{hp(AH_1T_1 \vee ... \vee AH_rT_r))}{hp(H_1T_1 \vee ... \vee H_rT_r))}\\
&= \frac{\sum_i hp(AH_iT_i)}{\sum_j hp(H_jT_j)}\\
&= \frac{\sum_i hp(A|H_iT_i)\cdot hp(H_iT_i)}{\sum_j hp(H_jT_j)}\\
&= \frac{\sum_i ch_{T_iH_i}(A)\cdot hp(H_iT_i)}{\sum_j hp(H_jT_j)}\\
&= \frac{\sum_i x\cdot hp(H_iT_i)}{\sum_j hp(H_jT_j)}\\
&= x.
\end{aligned}$$

The first four steps deserve further comment. I noted that if $\langle ch(A)=x\rangle E$ is admissible, then it can be expressed as a disjunction of the grounds of chance distributions, so if we wanted, we could use this fact to go directly from (9.9) to (9.11). The third step follows from the assumption that $\langle ch(A)=x\rangle E$ is admissible, and the definition of what this means given above. The fourth step replaces the disjunction $H_1T_1 \vee ... \vee H_qT_q$, where the terms are not mutually exclusive, with a disjunction of a subset of these terms, $H_1T_1 \vee ... \vee H_rT_r$, which are mutually exclusive. We can do this because if the grounding arguments H_iT_i and H_jT_j aren't mutually exclusive, then one will be a subset of the other. (T's are always mutually exclusive, and a history H is either different from history H', includes H', or is included in H'.)

9.4.2 $\mathbf{PP}_3 \Rightarrow \mathbf{PP}_2$

Suppose $ch_{TH}(A) = x$. Then $TH \Rightarrow \langle ch(A) = x\rangle$, and thus:

$$\begin{aligned} hp(A|TH) &= hp(A|\langle ch(A) = x\rangle TH) \\ &= x \\ &= ch_{TH}(A), \end{aligned} \tag{9.11}$$

The middle step requires $\langle ch(A) = x\rangle TH$ to be admissible. Using the definition of $\langle ch(A) = x\rangle E$ given above, we know this has to be the case. $\langle ch(A) = x\rangle E$ is admissible iff it can be expressed as a disjunction of TH terms. Since $TH \Rightarrow \langle ch(A) = x\rangle$, $\langle ch(A) = x\rangle \cap TH = TH$, which is a (trivial) disjunction of TH terms.

9.5 Deriving $\mathbf{PP}_1^-$ and $\mathbf{PP}_3^-$ from $\mathbf{PP}_2$

These derivations employ only L1-L4.

9.5.1 $\mathbf{PP}_2 \Rightarrow \mathbf{PP}_1^-$

Let $T_1H_1^t$ through $T_nH_n^t$ be all of the different chance theory and complete history up to t pairs such that the chance of A is x. By definition,

$$\langle ch_t(A) = x\rangle = T_1H_1^t \vee ... \vee T_nH_n^t \tag{9.12}$$

Then:

$$\begin{aligned} hp(A|\langle ch_t(A) = x\rangle) &= hp(A|T_1H_1^t \vee ... \vee T_nH_n^t) \\ &= \frac{hp(A \wedge (T_1H_1^t \vee ... \vee T_nH_n^t))}{hp(T_1H_1^t \vee ... \vee T_nH_n^t)} \\ &= \frac{\sum_i hp(AT_iH_i^t)}{\sum_j hp(T_jH_j^t)} \\ &= \frac{\sum_i hp(T_iH_i^t) \cdot hp(A|T_iH_i^t)}{\sum_j hp(T_jH_j^t)} \end{aligned} \tag{9.13}$$

$$
\begin{aligned}
&= \frac{\sum_i hp(T_i H_i^t) \cdot ch_{T_i H_i^t}(A)}{\sum_j hp(T_j H_j^t)} \\
&= x \cdot \frac{\sum_i hp(T_i H_i^t)}{\sum_j hp(T_j H_j^t)} \\
&= x.
\end{aligned}
$$

9.5.2 $\mathbf{PP_2 \Rightarrow PP_3^-}$

Let T_1H_1 through T_nH^n be all of the different chance theory and complete history pairs such that the chance of A is x. Then:

$$
\begin{aligned}
hp(A|\langle ch(A) = x\rangle) &= hp(A|T_1H_1 \vee ... \vee T_nH_n) \qquad (9.14) \\
&= hp(A|T_1H_1 \vee ... \vee T_mH_m) \\
&= \frac{hp(A \wedge (T_1H_1 \vee ... \vee T_mH_m))}{hp(T_1H_1 \vee ... \vee T_mH_m)} \\
&= \frac{\sum_i hp(AT_iH_i)}{\sum_j hp(T_jH_j)} \\
&= \frac{\sum_i hp(T_iH_i) \cdot hp(A|T_iH_i)}{\sum_j hp(T_jH_j)} \\
&= \frac{\sum_i hp(T_iH_i) \cdot ch_{T_iH_i}(A)}{\sum_j hp(T_jH_j)} \\
&= x \cdot \frac{\sum_i hp(T_iH_i)}{\sum_j hp(T_jH_j)} \\
&= x.
\end{aligned}
$$

The first step deserves further comment. The second step replaces the disjunction $H_1T_1 \vee ... \vee H_nT_n$, where the terms are not mutually exclusive, with a disjunction of a subset of these terms, $H_1T_1 \vee ... \vee H_mT_m$, which are mutually exclusive. We can do this because if two grounding arguments H_iT_i and H_jT_j aren't mutually exclusive, then one will be a subset of the other. (T's are always mutually exclusive, and a history H is either different from history H', includes H', or is included in H'.)

9.6 Hypothetical Priors, the Chance-Credence Relation, and Statistical Mechanics

If we formulate the chance-credence principle in terms of hypothetical priors, then we find that for chance theories like classical statistical mechanics, our priors only end up being constrained by trivial chances.

To see this, assume that (2.17) is our chance-credence principle. Now consider the Liouville measure of a state space S. If there are no particles in the systems of a state space, then the space will consist of a single point, and the associated chances will be trivial.[2] So let's confine our attention to state spaces whose systems have at least one particle. In classical mechanics there's no upper bound on the velocity of a particle, so the Liouville measure of any state space with particles will be infinite.

As before, assume the extended real number line and the standard extension of the arithmetical operators over it; in particular, that $\frac{x}{\infty} = 0$ if x is finite, and $\frac{\infty}{\infty}$ and $\frac{x}{0}$ are undefined. Now consider the chance of A relative to B, for some arbitrary propositions $A, B \subset S$. If $m(B) = \infty$ then $ch_{TB}(A)$ will either be undefined (if $m(A \cap B) = \infty$) or 0 (if $m(A \cap B) \neq \infty$). If $m(B) \neq \infty$, on the other hand, then $ch_{TB}(A)$ can take on non-trivial values. But if $m(B) \neq \infty$, then the chances require $hp(A|B)$ to be undefined, and (2.17) won't hook up our priors to these chances.

To see that the chances require $hp(A|B)$ to be undefined, suppose otherwise, i.e., suppose that $hp(B) > 0$. The chance of B relative to S will be

$$ch_{TS}(B) = \frac{m(B \cap S)}{m(S)} = 0, \tag{9.15}$$

since $m(B \cap S)$ is finite and m(S) infinite. And if $hp(B) > 0$ then $hp(S) > 0$,

[2] I follow Tolman (1979) here in not taking the total energy to be one of the relevant static properties. If we do adopt the total energy as one of these properties, then some of the details will be different.

since $B \subset S$, so $hp(B|S)$ is defined. Since both $hp(B|S)$ and $ch_{TS}(B)$ are well defined, (2.17) applies, and

$$\begin{aligned} ch_{TS}(B) &= 0 \\ hp(B|S) &= \\ \frac{hp(B \cap S)}{hp(S)} &= \\ \frac{hp(B)}{hp(S)} &= \\ &\Rightarrow hp(B) = 0, \end{aligned} \tag{9.16}$$

contradicting our supposition.

9.7 M4 for Multiple Layers

9.7.1 Formulating M4 for Multiple Layers

We can formulate M4 for multiple layers in the following way:

M4. Every chance theory T has the following structure:

(i) n layers of chance, L^1-L^n.

(ii) For each layer L^i, T can be partitioned into *coarse sets* C^i, each of which is a subset of the fine sets of the layer above it (if any).

(iii) For each layer L^i, the coarse sets can be partitioned into *fine sets* F^i.

(iv) Each coarse set C is associated with a countably additive measure m_{TC}, which is defined on an algebra that includes all of the fine sets of C but no proper subsets of these sets except the empty set.

(v) The chances of T are:

$$ch_{TB}(A) = \frac{m_{TC}(AB)}{m_{TC}(A)}, \tag{9.17}$$

where m_{TC} is the measure associated with the coarse set of the lowest layer that completely contains B.

9.8 Hypothetical Priors in Arbitrary Propositions

We can derive the (2.25) expression as follows.

We can partition the space of possible worlds into chance theories T_i, partition the chance theories into coarse sets C_j, partition the coarse sets into fine sets F_k, and partition the fine sets into individual worlds W_l. Now consider an arbitrary proposition, A. We know that if some sets X_i form a partition of A, we can express $hp(A)$ as

$$hp(A) = \sum_i hp(A \wedge X_i) = \sum_i hp(X_i)hp(A|X_i) \tag{9.18}$$

By applying (9.18) repeatedly for each of the above partitions, we can express $hp(A)$ as

$$\begin{aligned} hp(A) &= \sum_i hp(T_i)hp(A|T_i) && (9.19) \\ &= \sum_{i,j} hp(T_i)hp(C_j|T_i)hp(A|T_iC_j) \\ &= \sum_{i,j,k} hp(T_i)hp(C_j|T_i)hp(F_k|T_iC_j)hp(A|T_iC_jF_k) \\ &= \sum_{i,j,k,l} hp(T_i)hp(C_j|T_i)hp(F_k|T_iC_j)hp(W_l|T_iC_jF_k)hp(A|T_iC_jF_kW_l) \\ &= \sum_{i,j,k,l} hp(T_i)hp(C_j|T_i)hp(F_k|C_j)hp(W_l|F_k)hp(A|W_l) && (9.20) \end{aligned}$$

So we can determine the value of $hp(A)$ by figuring out the values of the five sets of terms in (9.20).[3]

(What if we adopt the tweaked version of M4 described in Appendix 9.7? How does the expression then turn out? If we express $hp(A)$ in terms of partitions of

[3] Again, I'm implicitly assuming that the indices i, j, k, l range over countably infinite members at most.

n layers, the expression becomes:

$$
\begin{aligned}
hp(A) \quad = \quad & \sum_{i,j,k,...,w,x,y,z} hp(T_i)hp(C_j^1|T_i)hp(F_k^1|C_j^1) \; ... \\
& ... \; hp(C_x^n|F_w^{n-1})hp(F_y^n|C_x^n)hp(W_z|F_y^n)hp(A|W_z)
\end{aligned} \tag{9.21}
$$

Since different chance theories will have different numbers of layers, the completely general expression will need to include as many layers are any possible chance theory has. Theories with fewer layers will then be treated as theories with his many layers, with the superfluous layers having a single coarse set and fine set, these sets being identical to the fine set of the last "real" layer containing them. (This will get tricky for theories with an infinite numbers of layers, of course. As with most infinity complications, I'm putting this worry aside.))

9.9 Every Well-Defined Conditional Chance Corresponds to an Unconditional Chance

Another consequence of M1-M4 is that every well-defined conditional chance $ch_{TB}(A|E)$ will have a corresponding unconditional chance $ch_{TBE}(A)$. So:

$$
\begin{aligned}
ch_{TB}(A|E) \quad &= \quad \frac{ch_{TB}(AE)}{ch_{TB}(E)} \\
&= \quad \frac{\left(\frac{m(AEB)}{m(B)}\right)}{\left(\frac{m(EB)}{m(B)}\right)} \\
&= \quad \frac{m(AEB)}{m(BE)} \\
&= \quad ch_{TBE}(A)
\end{aligned} \tag{9.22}
$$

The key is the last step: how can we be sure that $ch_{TBE}(A)$ will always be well-defined? As noted earlier, we know it will be well-defined iff three conditions are satisfied: (a) T is a complete chance theory and BE is a subset of a coarse set C of T, (b) the ratio of $m(ABE)$ to $m(BE)$ is defined, and (c) ABE and BE are elements of S, the algebra over which m is defined.

Because $ch_{TB}(A|E)$ is well-defined, we know that T is a complete chance theory and that B is a subset of some coarse set C of T. Since BE is a subset of B, it follows that BE is a subset of C as well. So (a) is satisfied. The ratio of $m(ABE)$ to $m(BE)$ will be well-defined iff (i) $m(BE) \neq 0$, (ii) both $m(ABE)$ to $m(BE)$ aren't both infinite, and (iii) both $m(ABE)$ to $m(BE)$ are well-defined. (As before, I'm assuming the extended real number line.) Since all three of these conditions must be satisfied in order for $ch_{TB}(A|E)$ to be well-defined, (b) is satisfied. Finally, we know that both ABE and BE are elements of S, since $m(ABE)$ and $m(BE)$ are well-defined. So (c) is satisfied. Since (a), (b) and (c) are satisfied, $ch_{TBE}(A)$ will be well-defined.

9.10 Hitchcock's Dutch Book Argument

Hitchcock (2004) considers the following betting situation: a bookie undergoes the experiment with Sleeping Beauty, and offers her a bet on Sunday night, as well as every time they wake up. He then shows that if Beauty takes $\$\frac{1}{2}$ to be a fair price for a bet that pays \$1 if heads comes up, then the bookie can construct a Dutch book against her. More generally, one can show that if Beauty takes anything other than $\$\frac{1}{3}$ to be a fair price for a \$1 bet when she wakes up, she can be Dutch booked.

Let us adopt the following notation for bets. Bets will be represented by five letters and a subscript and superscript. The first letter will be B (for bet), the next three letters will indicate the day on which the bet is offered and accepted, and the fifth letter will indicate what is being bet on, the subscript will indicate the amount paid for the bet, and the superscript will indicate the payoff if the centered proposition turns out to be true. (Occassionally I will omit the superscript when the payoff of the bet is the standard \$1.) So, for example, 'B-SUN-$\text{T}^{\$1}_{\$1/2}$' will stand for a Sunday bet on tails that was bought for $\$\frac{1}{2}$ and pays \$1 if heads.

'B-MON-H$^{\$3}_{\$1}$' will be a Monday bet on heads that pays \$3 if heads comes up, and that was bought for \$1.

The Dutch books Beauty is vulnerable to can be divided into two kinds of cases: (i) those where after waking up she takes $x > \frac{1}{3}$ to be the proportion of the payoff that a bet on heads is worth, and (ii) those where after waking up she takes $x < \frac{1}{3}$ to be the proportion of the payoff that a bet on heads is worth. (I assume, as usual, that if she takes x to be a fair proportion of the payoff to pay for a bet on heads, then she'll take $1 - x$ to be a fair proportion of the payoff to pay for a bet on tails.) Let's look at these cases in order.

9.10.1 $x > \frac{1}{3}$

In this case the bookie will offer Beauty the following bets: (i) B-SUN-T$^{\$(1+x)}_{\$1/2(1+x)}$, (ii) B-MON-H$^{\$1}_{\$x}$ and (iii) B-TUE-H$^{\$1}_{\$x}$ (if they wake up Tuesday). So the net gain or loss of each of these bets, if heads or tails comes up, will be:

	H	T
B-SUN-T$^{\$(1+x)}_{\$1/2(1+x)}$	$-\frac{1}{2} \cdot (1+x)$	$\frac{1}{2} \cdot (1+x)$
B-MON-H$^{\$1}_{\$x}$	1-x	-x
B-TUE-H$^{\$1}_{\$x}$		-x
Net Gain/Loss:	$\frac{1}{2} - \frac{3x}{2}$	$\frac{1}{2} - \frac{3x}{2}$

Since $x > \frac{1}{3}$, $\frac{1}{2} - \frac{3x}{2} < 0$, so Beauty will lose money no matter what.

9.10.2 $x < \frac{1}{3}$

In this case the bookie will offer Beauty the opposite sides of the above bets: (i) B-SUN-H$^{\$(1+x)}_{\$1/2(1+x)}$, (ii) B-MON-T$^{\$1}_{\$(1-x)}$ and (iii) B-TUE-T$^{\$1}_{\$(1-x)}$ (if they wake up

Tuesday). So the net gain or loss of each of these bets, if heads or tails comes up, will be:

	H	T
B-SUN-H$^{\$(1+x)}_{\$1/2(1+x)}$	$\frac{1}{2} \cdot (1+x)$	$-\frac{1}{2} \cdot (1+x)$
B-MON-T$^{\$1}_{\$(1-x)}$	-(1-x)	x
B-TUE-T$^{\$1}_{\$(1-x)}$		x
Net Gain/Loss:	$-\frac{1}{2} + \frac{3x}{2}$	$-\frac{1}{2} + \frac{3x}{2}$

Since $x < \frac{1}{3}$, $-\frac{1}{2} + \frac{3x}{2} < 0$, so Beauty will lose money no matter what.

9.11 Sleeping Beauty and Decision Theory

Given that her credences in heads/tails are $y/1-y$, what should Beauty consider a fair price for a \$1 bet on heads or tails be when she wakes up on Monday morning? As it turns out, this varies depending on the kind of decision theory one adopts. Let's first consider evidential decision theory and a standard version of causal decision theory, in turn.

9.11.1 Fair Bets According to Evidential Decision Theory

Given evidential decision theory, what's a fair price for a \$1 bet on heads when Beauty wakes up, given that her credence in heads is y? Assume, as is usually implicit, that she is certain that she will bet the same way on both days (if there is a second day), and that her utility is linear in dollars. Let 'B-NOW-H$^{\$1}_{\$x}$' be the centered proposition that she's at a world where she pays \$$x$ for such a bet, and that her temporal location is either Monday or Tuesday. To find the fair price of the bet, we need to find the value of x which makes the evidential expected utility of B-NOW-H$^{\$1}_{\$x}$ the same as the evidential expected utility of not accepting the bet. Since if you don't accept the bet then your other temporal counterpart (if any) won't accept a bet either, we know that the evidential expected utility of not accepting the bet will be 0. So we need to find the value of x which makes the evidential expected utility of B-NOW-H$^{\$1}_{\$x}$ 0.

Her current credences, before she makes a decision, will be split between four salient possibilities: the coin lands heads and she does/doesn't accept the bet, and the coin lands tails and she does/doesn't accept the bet. (We don't need to worry about the possibility of her accepting a bet on Monday and not Tuesday, or vice versa, since she knows that Monday and Tuesday counterparts will bet the same way.) Her credence in in the two heads possibilities is y, and her credence in the two tails possibilities is $1 - y$. Presumably her credence that she'll accept or decline the bet is independent of the outcome of the coin toss; let her credence that she'll accept be z, and her credence that she'll decline be $1 - z$. So her pre-decision credences in the four possibilities will be: yz that the coin lands heads and she bets, $y(1 - z)$ that the coin lands heads and she doesn't bet, $(1 - y)z$ that the coin lands tails and she bets, and $(1 - y)(1 - z)$ that the coin lands tails and she doesn't bet.

Her credence in B-NOW-H$^{\$1}_{\$x}$ will be the sum of first and third of these possibilities, or $yz + (1 - y)z = z$. Given that she accepts this bet, there are only two centered propositions about the outcome of the coin toss and her waking-up betting behavior that she has a non-zero credence in: H$\wedge$MON$\wedge$B-MON-H$^{\$1}_{\$x}$ and T$\wedge$(MON$\vee$TUE)$\wedge$B-MON-H$^{\$1}_{x}\wedge$B-TUE-H$^{\$1}_{\$x}$. And her pre-decision credences in these possibilities will be yz and $(1 - y)z$, respectively.

With this in hand, we can now calculate the evidential expected utility of B-NOW-H$^{\$1}_{\$x}$:

$$\begin{aligned} EEU(\text{B-NOW-H}^{\$1}_{\$x}) &= cr(H \wedge \text{MON} \wedge \text{B-MON-H}^{\$1}_{\$x}|\text{B-NOW-H}^{\$1}_{\$x}) \qquad (9.23) \\ &\cdot\ u(H \wedge \text{MON} \wedge \text{B-MON-H}^{\$1}_{\$x}) \end{aligned}$$

$$\begin{aligned}
&+ \quad cr(T \wedge (\text{MON} \vee \text{TUE}) \wedge \text{B-MON-H}^{\$1}_{\$x} \\
&\qquad \wedge \text{B-TUE-H}^{\$1}_{\$x} | \text{B-NOW-H}^{\$1}_{\$x}) \\
&\cdot \quad u(T \wedge (\text{MON} \vee \text{TUE}) \wedge \text{B-MON-H}^{\$1}_{\$x} \wedge \text{B-TUE-H}^{\$1}_{\$x}) \\
&= \quad y \cdot (1-x) + (1-y) \cdot (-2x) \\
&= \quad y + xy - 2x
\end{aligned}$$

Setting this equal to 0 and solving for x gives us:

$$x = \frac{y}{2-y}, \tag{9.24}$$

which is our answer. Alternatively, if we want to find out what credence y in heads she must have if she's a rational agent and takes x to be the fair price for a \$1 bet on heads, then we can solve for y:

$$y = \frac{2x}{1+x}. \tag{9.25}$$

So if her credence in heads/tails is $\frac{1}{2}/\frac{1}{2}$ when she wakes up, then she should take a \$1 bet on heads to be worth $\$\frac{1}{3}$. On the other hand, if her credence in heads/tails is $\frac{1}{3}/\frac{2}{3}$ when she wakes up, then she should take a \$1 bet on heads to be worth $\$\frac{1}{5}$.[4]

Given the Dutch book results we saw above, we can see that if Beauty is an evidential decision theorist, she will be vulnerable to a Dutch book unless her credence in heads/tails when she wakes up is $\frac{1}{2}/\frac{1}{2}$.

9.11.2 Fair Bets According to Causal Decision Theory

A standard way to cash out causal decision theory is to replace the conditional probabilities $p(A|B)$ used by the evidential decision theorist with imaging probabilities $p(A||B)$. Since imaging relies on a specification of similarity relations, this

[4] I.e., what you'd expect for 4:1 odds.

characterization is very flexible. By playing around with the similarity relations, we can get radically different kinds of imaging functions.

The causal decision theorist usually wants the imaging function $p(A\|B)$ to capture something like the probability of A coming about if B were to be true, and everything causally independent of B were the same as it actually is. Let's start by assuming that the imaging function captures something like this.

Again, let's ask what Beauty should consider a fair price x for a \$1 bet on heads when she wakes up, given that her credence in heads is y. As before, we want to find the value of x which makes the causal expected utility of B-NOW-$\mathrm{H}^{\$1}_{\$x}$ be the same as the causal expected utility of not accepting the bet. This time, however, things are a bit trickier.

First, we have to consider a wider range of possibilities than before. In the evidential case, the fact that there are only two propositions about the outcome of the coin toss and her waking-up betting behavior that she has a non-zero credence in (i.e., H∧MON∧B-MON-$\mathrm{H}^{\$1}_{\$x}$ and T∧(MON∨TUE)∧B-MON-$\mathrm{H}^{\$1}_{\$x}$∧B-TUE-$\mathrm{H}^{\$1}_{\$x}$) allowed us to ignore the other possibilities when we calculated her EEU. But in this case we can't do that: we also have to consider the propositions T∧(MON∨TUE)∧B-MON-$\mathrm{H}^{\$1}_{\$x}$∧TUE(no bet) and T∧(MON∨TUE)∧MON(no bet)∧B-TUE-$\mathrm{H}^{\$1}_{\$x}$. This is because the way Beauty decides to bet on Monday is (presumably) causally independent of how she bets on Tuesday. She won't bet how she does on Tuesday *because* she bet that way on Monday; rather, she'll bet the same way on Monday and on Tuesday because her dispositions to bet happen to be exactly the same on both days. And when we evaluate the probability of one of these propositions coming about assuming we act in a given way, but keeping everything causally independent of our act the same, we have to allow for the possibility that our other temporal counterpart (if any) *doesn't* act the same as she does, since how Beauty's two temporal parts act is causally independent in the relevant sense.

The second reason things get trickier is that it won't generally be the case that the causal expected utility of not accepting a bet will be 0. This is so for some of the reasons just considered. If the coin lands heads, then we won't be the only one who the bookie offers the bet—he'll also offer it to our temporal counterpart on the other day. And even if we decline to accept the bet on heads, the counterpart might accept. If he does, we'll lose money, since the coin landed tails. So given that we refuse the bet, we have no possibility of making money, and some possibility of losing money. So the next causal expected utility can be less than 0. That means we have to explicitly work out the causal expected utility of not accepting the bet as well, in order to figure out what value of x is a fair price for the bet on heads.

Finally, this means we have to consider two more propositions about the waking-up and betting behavior of ourselves and our temporal counterparts that we could ignore if we were calculating the causal expected utility of performing the act alone—H∧MON∧MON(no bet) and T∧(MON∨TUE)∧MON(no bet)∧TUE(no bet)—which both *are* possible if we refuse to accept the bet offered to us.

So, there are six centered propositions to consider. With a little work (to be expounded upon further, below) we can figure out (i) the probability of these centered propositions given that we've imaged on accepting the bet or refusing the bet, and (ii) the monetary gain or loss given each of these centered propositions, and thus (iii) the causal expected utility contribution from each of these centered propositions (for convenience, I've abbreviated the names of the centered propositions):

centered proposition	utility	cr imaged on bet	CEU (bet)	cr imaged on no bet	CEU (no bet)
H+BM	1-x	y	y-yx	0	0
H+noBM	0	0	0	y	0
T+BM+BTu	-2x	2b	-4xb	0	0
T+BM+noBTu	-x	a	-xa	b	-xb
T+noBM+BTu	-x	a	-xa	b	-xb
T+noBM+noBTu	0	0	0	2a	0
Total		y+2b+2a = 1	y-yx-4xb-2xa	y+2b+2a = 1	-2xb

The utilities associated with each centered proposition are self-explanatory. Given the credences, the CEUs are self-explanatory as well: they're just the products of the utilities and the imaged credences of the corresponding centered propositions. But the credences are less straightforward. Let's go through them.

Beauty's credence in heads/tails before imaging is $y/1-y$, by stipulation. Now, imaging on a centered proposition moves the credence assigned to possibilities incompatible with that centered proposition over to the nearest possibility that is compatible with that centered proposition. Neither of the centered propositions being imaged on eliminates all of the heads or tails possibilities, so given the kinds of similarity relations we're considering, the nearest possibilities are presumably ones where the outcome of the coin toss is the same. So Beauty's credence in heads/tails after imaging will remain the same: $y/1-y$.

The centered proposition that the coin lands heads and the Monday bet is declined is ruled out completely once we image on accepting the bet. (If the coin lands heads and we accept the bet, then it can't be the case that the Monday bet is declined.) So all of the credence assigned to heads—y—goes to H+BM after we image on accepting the bet. Likewise, the centered proposition that the coin lands heads and the Monday bet is accepted is ruled out completely once we image on declining the bet. So all of the credence assigned to heads goes to H+noBM after we image on declining the bet.

Similar considerations rule out the possibility of the coin landing tails and both the Monday and Tuesday bets being declined, once we image on accepting the bet. So T+noBM+noBTu will get no credence after we image on accepting the bet. And, likewise, the possibility of the coin landing tails and both the Monday and Tuesday bets being accepted is ruled out once we image on accepting the bet, so T+BM+BTu will get no credence after we image on declining the bet.

Now, how does the credence assigned to tails, $1 - y$, get divided up between the three surviving tails centered propositions in each case? As it turns out, this is underdetermined. But we can say a number of things about these credence assignments.

Symmetry considerations indicate Beauty's credences in T+BM+noBTu and T+ noBM+BTu will be the same, since she has no way to distinguish between Monday and Tuesday. (This isn't an entirely innocuous assumption, but given our concerns, there's no harm in going along with the standard symmetric assignments to Monday and Tuesday given tails.) Let a be the credence these centered propositions receive after imaging on accepting the bet, and b be the credence they receive after imaging on declining the bet.

Causal considerations provide another constraint on these credence assignments. According to standard causal decision theory, the similarity constraints which characterize the imaging function are such that $p(A\|B) = p(A\|\neg B)$, if A is causally independent of B. Now consider the bet that your temporal counterpart (if any) makes. This is causally independent of how you bet. So your credence that your temporal counterpart accepts the bet should be the same regardless of whether you bet or not. If you accept the bet then your counterpart's acceptance means T+BM+BTu is true, if you decline the bet then your counterpart's acceptance means either T+BM+noBTu or T+noBM+BTu are true. Since these should be the same, it follows that your credence in T+BM+BTu after imaging on accepting the bet must be equal to the sum of your credence in T+BM+noBTu and T+noBM+BTu after imaging on declining the bet. I.e., it follows that your credence in T+BM+BTu after imaging on accepting the bet must be $2b$.

Likewise, your credence that your temporal counterpart declines the bet should be the same whether you bet or not. A similar train of reasoning allows us to deduce that our credence in T+noBM+noBTu after imaging on declining the bet must be equal to the sum of your credence in T+BM+noBTu and T+noBM+BTu after accepting the bet, i.e., $2a$.

This explains the credence assignments listed on the chart. Now, back to our question: what should Beauty consider to be a fair price x for a bet on heads? To get the fair price, we need to find the value of x such that the causal expected utility of accepting the bet is the same as the causal expected utility of declining the bet:

$$y - yx - 4xb - 2xa = -2xb \tag{9.26}$$

Adding $2xb$ to both sides we can leap into the algebraic fray:

$$0 = y - yx - 2xb - 2xa = y(1-x) - x(2b+2a) \tag{9.27}$$

We know that $y + 2a + 2b = 1$ (since Beauty's post-imaging credences must sum to 1), so we can replace $2a + 2b$ with $1 - y$:

$$y(1-x) - x(2b+2a) = y(1-x) - x(1-y) \Rightarrow \frac{y}{(1-y)} = \frac{x}{(1-x)} \Rightarrow x = y \tag{9.28}$$

At long last, we have our answer.

So if Beauty is a causal decision theorist, and her credence in heads/tails is $\frac{1}{2}/\frac{1}{2}$ when she wakes up, then she should take a \$1 bet on heads to be worth $\$\frac{1}{2}$. If, on the other hand, her credence in heads/tails is $\frac{1}{3}/\frac{2}{3}$ when she wakes up, then she should take a \$1 bet on heads to be worth $\$\frac{1}{3}$.

Given the Dutch book results we saw above, we can see that if Beauty is a causal decision theorist, she will be vulnerable to a Dutch book unless her credence in heads/tails when she wakes up is $\frac{1}{3}/\frac{2}{3}$.

9.12 Halpern's Rules

Halpern (2005) has suggested a way of capturing Elga's response, and Halpern and Tuttle (1993) propose an account of temporal belief change. Both of these proposals are only intended to accommodate beliefs about temporal location, not self-locating beliefs in general. So in this context, we should understand centered worlds as just world-time pairs, and centered propositions as sets of these pairs. Let's look at each of these rules in turn.

Halpern (2005) proposes a way of capturing Elga's response. Given an appropriate translation, Halpern's Elga rule (HER) can be expressed as follows. Consider a subject whose current evidence is the centered proposition e. According to HER, her new credence in a centered proposition a, $cr_e(a)$, should be:[5]

$$cr_e(a) \overset{\text{def}}{=} \sum_{(i|c_i \in a)} \left(\frac{hp(\overline{c_i e})}{\sum_j hp(W_j \overline{e}) \cdot m_j(e)} \right) \tag{9.30}$$

[5] Halpern (2005) actually presents Elga's rule as (after appropriate translation):

$$cr_e(c) = \frac{hp(\overline{ce})}{\sum_i hp(W_i \overline{e}) \cdot \#(W_i e)} \tag{9.29}$$

where $\#(\cdot)$ is a function that spits out the number of centered worlds contained in the centered proposition. By summing over all of the centered worlds in a centered proposition, replacing the $\#(\cdot)$ expression with a $m_i(\cdot)$ expression, and we get the expression for one's credences in centered propositions given above.

Like CeC, this is a hypothetical prior rule. But HER differs from CeC in two ways. (Three, if you include the fact that HER only applies to cases of temporal location.) First, the hypothetical prior function is only defined over worlds, not centered worlds. Second, Elga's Indifference Principle is built directly into HER.

How does HER compare to CeC_E? Although the two accounts yield the same answers in canonical cases like the sleeping beauty case, the two accounts can come apart. We know this because CeC_E is ambiguous, while HER is not: the credences CeC_E assigns will depend on how the Continuity Principle is precisified.

This difference may appear to be an advantage for HER. Because of this determinateness, HER may appear to circumvent the continuity problems I raise in section 5.8.2. But matters are more complicated: this is a much a problem for HER as an advantage.

Like CeC, HER is a hypothetical priors rule. But unlike CeC, HER is incompatible with the standard characterizations of hypothetical priors. If you apply the rule to generate the credences of a subject with no evidence ($e = \Omega$), this rule will not usually yield hp again. So if we adopt HER, we can't understand hypothetical priors as one's initial credences before they got any evidence, or what one's credences should be if they had no evidence, or anything like that. This means that none of the three ways of understanding hypothetical priors described in section 1.2 will do. And this leaves us without a grasp of what hypothetical priors are, and *a fortiriori*, without a grasp of HER itself.

The best route for the proponent of HER is to adopt something like the third understanding of hypothetical priors sketched in section 1.2—as a merely functional device—but without the claim that they correspond to something like "the credences one ought to have if they had no evidence". This avoids the immediate problem that we raised above, but leaves us without a tangible grasp on what priors are supposed to intuitively correspond to. (Though perhaps this shouldn't be a concern on a functionalist understanding.)

Halpern and Tuttle (1993) propose an account of updating that accommodates temporal belief similar to CoC_M. Given an appropriate translation, Halpern and Tuttle's rule (HTR) can be expressed as follows. Consider a subject whose current evidence is the centered proposition e. According to HTR, her new credence in a *de dicto* proposition A, $cr_e(A)$, should be:[6]

$$cr_e(A) = hp(A|\bar{e}) \tag{9.32}$$

Like CoC, this is a hypothetical prior rule. But HTR differs from CeC in two ways. (Three, if you include the fact that HTR only applies to cases of temporal location.) First, the hypothetical prior function is only defined over worlds, not centered worlds. Second, HTR is silent about how a subject's credence in a world should be divided up among her doxastic alternatives at that world.

How does HTR compare to CoC_M? The two accounts agree on how a subject should assign credences to worlds. But HTR is silent about how to assign credences to centered worlds. So CoC (and *a fortiriori* CoC_M) can be seen as a precisification of HTR. HTR doesn't appear to have any of the continuity worries I describe in section 5.10. But this can't be thought of as an advantage, since with respect to worlds CoC doesn't have any continuity worries either. Continuity worries only arise when we consider how to assign credences to centered worlds, and HTR is silent about that.

Of course, it's relatively easy to modify HTR so that it does assign credences to centered worlds in an unambiguous manner. Namely, we can effectively built

[6] Halpern and Tuttle (1993) actually presents the rule (appropriately translated) as:

$$cr_e(eW) = hp(W|\bar{e}) \tag{9.31}$$

But assuming you assign a credence of 0 to possibilities incompatible with your evidence, your credence in eW will be equal to your credence in W. So we can express one's credence in any *de dicto* proposition A in the manner given above.

Elga's Indifference Principle into the rule (call it HTR*) as follows. Consider a subject whose current evidence is the centered proposition e. According to HTR*, her new credence in a centered proposition a, $cr_e(a)$, should be:

$$cr_e(a) \overset{\text{def}}{=} \sum_{(i|c_i \in a)} \frac{hp(\overline{c_i e})}{m_{\overline{c_i}}(e)} \tag{9.33}$$

HTR* yields the same credence assignments as CoC_M. And unlike CoC, HTR* doesn't appear to have any continuity problems. But HTR* has the same problems making sense of hypothetical priors as HER does. So it's unclear which we should prefer.

9.13 Evaluating Elga's Indifference Principle

I formulated Elga's Indifference Principle as the following constraint:

$$cr_e(\cdot|W_i) = m_i(\cdot|e), \text{ if } cr_e(\cdot|W_i) \text{ is defined.} \tag{9.34}$$

What should we think of this principle?

I'm generally inclined to be suspicious of indifference principles. As far as indifference principles go, though, I take Elga's Indifference Principle to be reasonable. It's relatively plausible that if we have several alternatives at a world, we should have equal credence in each. It's less clear whether (9.34) is plausible in cases where you have an infinite number of alternatives* at some world W_i. (As I mentioned in section 4.1, it will be convenient to use Lewis's other characterization of doxastic worlds and alternatives here, and I will mark these uses with an asterix.) In these cases, (9.34) sets our credences in them in accordance with the canonical self-locating measure of W_i, m_i. But since we haven't said anything more about what m_i is, it's hard to tell whether these assignments are intuitive or not. Saying more about what m_i is like in the infinite case is an interesting question. Indeed, I take this to be the biggest challenge facing a complete description of Elga's Indifference Principle. I don't have anything interesting to say about this project, however, so I'll just note it, and move on.

Weatherson (2005) has offered several further criticisms of Elga's Indifference Principle. These criticisms divide up into roughly four categories:

1. Elga's Indifference Principle blurs the division between risk with uncertainty.

2. Elga's Indifference Principle, and Elga's picture of evidence in general, conflicts with externalist theories of experience, externalist theories of justification, and some mainstream accounts of vagueness.

3. In cases where a subject has an infinite number of alternatives*, Elga's Indifference Principle either requires the rejection of countable additivity, or places no real constraint on our credences.

4. Elga's Indifference Principle won't yield the desired credence constraints in cases where subject's have an infinite number of doxastic worlds*.

I think none of these criticisms are deep problems for Elga's account. Let me spell out why.

The first criticism assumes that there are two different ways in which a subject can be unsure about something. One of these ways of being unsure—"Risk"—can be adequately represented by something like degrees of belief. The other way of being unsure–"Uncertainty"—is better captured by employing something like sets of Risk-encoding probability functions which you're Uncertain between. Given these assumptions, Weatherson argues that Elga's principle conflates the two ways of being unsure. It applies to cases of Uncertainty, but it treats it like a case of Risk (it assigns precise degrees of belief).

I'm not sure I understand the difference between Risk and Uncertainty, but even putting that aside, this criticism seems misplaced. Like us, Elga is assuming the standard idealizations of the Bayesian framework: A1, A2, and the like. Weatherson's criticism is essentially that one of these idealizations—that a subject's epistemic state at a time can be represented by a (single) probability function (A1)—is false. But this isn't an interesting criticism of Elga's principle, just a rejection of the assumptions Elga starts with. Since, like Elga, we are adopting the standard Bayesian idealizations, we can put this criticism aside.

The second criticism is similar: Weatherson notes that Elga's discussion assumes a picture of evidence and justification that conflicts with externalist theories of experience, externalist theories of justification, and some mainstream accounts of vagueness. I've addressed something like this criticism in section 4.2. In any case, the remarks about the previous criticism apply here as well. Since we're assuming a picture of evidence similar to Elga's (A4-A6), we can put this criticism aside.

For his third criticism, Weatherson points out that if we understand Elga's principle as he formulates it—alternatives* at the same world should be assigned equal credences—then we run into difficulties in cases where infinite numbers of alternatives* at a world are compatible with our evidence. If there are a countably infinite number of them, then we cannot assign them all equal credences without violating countable additivity. If there are an uncountably infinite number of them, then the only way to satisfy this constraint is to assign them all a 0 credence. But this places very little constraint on our credences, since this constraint is compatible with almost any probability measure.

This criticism is correct, but easily overcome. If we adopt the formulation of Elga's Indifference Principle given above (9.34) then neither of these problems arise.[7]

[7] How would one treat something like CeC$_E$'s Continuity Principle in the infinite case? Using the terminology defined below, here's one way to proceed: Instead of having the temporal

The fourth criticism also raises infinity worries. Weatherson points out that if a subject has an infinite number of doxastic worlds*, then her credence in each of these worlds is likely to be 0. But then the prescription to divide the credence assigned to a world evenly among the alternatives* at that world places virtually no constraint on our credences.

Getting around this requires a bit more work, but I don't see any principled reason to think one can't extend (9.34) to such cases. Let me lay how one might do so using a simple model.

Assume we can form a bijection between worlds and a segment of the real number line, and that we can form a bijection between the centered worlds at each world and a segment of the real number line. Line up the worlds with some segment of the real number line $S1$ (which I'll use the variable x to range over). Line up the centered worlds with some segment of the real number line $S2$ (which I'll use the variable y to range over), such that $\forall i, j(m_i([a,b]) = m_j([a,b]))$.

Then we can represent our credence function as a probability density ρ of two variables, x and y. To find our credence in a given a centered proposition that's a Borel set of $S1 \times S2$, we can just integrate over the area(s) of $S1 \times S2$ containing the centered worlds corresponding to that centered proposition.[8] So, for example,

successor relation defined for pairs of worlds, define it for centered propositions of the form $q = \{(x, y)|x = m, y \in [a, b]\}$ (segments of $S2$ at a given world). If the relation holds between such segments, then they're continuous. And in the case of CeC, we can interpret that as the following priors constraint: given any two centered propositions of the form q, a and b, that are in the same subjective state, the Continuity Principle requires that the ratio between a subject's priors in a and b be the same as the ratio between her priors in any segment that's a temporal successor of a and any segment that's a temporal successor of b. (This needs further expansion in order to address the second kind of infinity worry Weatherson raises, of course.)

[8] There's no guarantee that we will be able to produce a well-defined credence for centered propositions which aren't Borel sets of $S1 \times S2$. But this shouldn't surprise us: if the space of possibilities is isomorphic to the reals, for example, then (given the axiom of choice) we know that there will be unmeasurable sets of these possibilities, and thus sets of possibilities we won't have well-defined credences for. See Hajek (2003) for a discussion of some related issues.

our credence in the centered proposition $q = \{(x,y)|x \in [m,n], y \in [a,b]\}$ will be:

$$cr(q) = \int_m^n \int_a^b \rho(x,y). \tag{9.35}$$

Now we can impose the indifference constraint on $\rho(x,y)$ in this case in the following way. Consider a subject whose evidence is e, and consider two centered propositions $q = \{(x,y)|x \in [m,n], y \in [a,b]\}$ and $r = \{(x,y)|x \in [m,n], y \in [c,d]\}$, which are subsets of e. Then for any i, m and n, we require ρ to be such that:

$$\frac{m_i([a,b])}{m_i([c,d])} = \frac{\int_m^n \int_a^b \rho(x,y)}{\int_m^n \int_c^d \rho(x,y)}. \tag{9.36}$$

9.14 The Many Brains Argument

For simplicity, assume that there are only two worlds under consideration, one normal world and one brain-duplicating world; it's easy to see how the result generalizes to multiple worlds. Let S be the stable world, and D be the duplicating world.

Consider the alternatives focused on the original (non-brain) individual at the S and D worlds. As time changes you will replace these alternatives with new alternatives at those worlds, centered on the same individual and a later time. (At the D world, of course, you will also be replacing old brain-centered alternatives with their temporal successors, as well as adding entirely new brain alternatives.) The new non-brain alternatives and the old non-brain alternatives they replaced will be continuous according to CeC$_E$'s Continuity Principle. We saw in section 5.3 that given centered conditionalization, the Continuity Principle requires that the ratios of priors between new and old continuous alternatives be the same. So the ratio of your priors in the non-brain alternatives at the D and S worlds at a time will be constant. That is, if we let pr_{S_t} and pr_{D_t} be your priors in the non-brain alternatives at the D and S worlds at t, the Continuity Principle entails that $\forall t \left(\frac{\text{pr}_{D_t}}{\text{pr}_{S_t}} = k\right)$, for some constant k.

Elga's Indifference Principle entails that one's credences in alternatives at a world be the same, and thus (given centered conditionalization) that one's priors in alternatives at a world be the same. So one's prior in the brain alternatives centered on world D and time t will be the same as your prior in the non-brain alternative centered on world D and time t, pr_{D_t}.

Now, let N_{W_t} be the number of alternatives you have at time t that are centered on a world W, and let $\text{cr}_t(W)$ be your credence at t in W. Assume the brains are created one at a time, and choose temporal units and a temporal origin such that (a) $\text{N}_{D_0} = \text{N}_{S_0} = 1$, and (b) $\text{N}_{D_t} = t{+}1$. Since you only ever have one alternative centered on S, $\forall t$ ($\text{N}_{S_t} = 1$).

Centered conditionalization and the above then entail that:

$$\begin{aligned} \text{cr}_t(D) &= \frac{N_{D_t} \cdot \text{pr}_{D_t}}{N_{D_t} \cdot \text{pr}_{D_t} + N_{S_t} \cdot \text{pr}_{S_t}} \\ &= \frac{N_{D_t} \cdot \text{pr}_{D_t}}{N_{D_t} \cdot \text{pr}_{D_t} + N_{S_t} \cdot \frac{\text{pr}_{D_t}}{k}} \\ &= \frac{N_{D_t}}{N_{D_t} + \frac{N_{S_t}}{k}} \\ &= \frac{t+1}{t+1+\frac{1}{k}}. \end{aligned} \tag{9.37}$$

Thus:

$$\lim_{t\to\infty} (\text{cr}_t(D)) = \lim_{t\to\infty} \left(\frac{t+1}{t+1+\frac{1}{k}}\right) = 1. \tag{9.38}$$

9.15 The Sadistic Scientist's Argument

Again, for simplicity assume that there are only two worlds under consideration, one normal world and one brain-duplicating-and-destroying world. Let S be the stable world, and D be the duplicating-and-destroying world.

As before, let N_{W_t} be the number of alternatives you have at time t that are centered on a world W, and let $\mathrm{cr}_t(W)$ be your credence at t in W. Choose temporal units and a temporal origin such that if $t < 0$ or $t > n$, then $N_{D_t} = 1$, and if $0 \leq t \leq n$, then $N_{D_t} = (n+1) - t$. (So n is the number of brains that will be created in D at time $t = 0$, and one of these brains will be destroyed every unit of time thereafter.)

As before, let pr_{S_t} and pr_{D_t} be your priors in the non-brain alternatives at the D and S worlds at t. Now consider the alternatives focused on the original (non-brain) individual at the S and D worlds. As time changes you will replace these alternatives with new alternatives at those worlds. (At the D world, of course, you will also be replacing old brain-centered alternatives with their temporal continuants, as well as adding entirely new brain alternatives.) The new non-brain alternatives and the old non-brain alternatives they replaced will satisfy the condition for continuity according to CeC_L's Continuity Principle until time $t = 0$, when the brains are created. So the Continuity Principle entails that for $t < 0$, $\left(\frac{\mathrm{pr}_{D_t}}{\mathrm{pr}_{S_t}} = k\right)$, for some constant k. The conditions for continuity will also hold after the brains are created, so the Continuity Principle entails that for $t \geq 0$, $\left(\frac{\mathrm{pr}_{D_t}}{\mathrm{pr}_{S_t}} = l\right)$, for some constant l.

Elga's Indifference Principle entails that one's credences in alternatives at a world be the same, and thus (given centered conditionalization) that one's priors in alternatives at a world be the same. So one's prior in the brain alternatives at D at t will be the same as your prior in the non-brain alternative at D, pr_{D_t}. The No-Increase Principle entails that your credence in D shouldn't change when the new brains are created at $t = 0$. This, centered conditionalization and the above entail that $l = k/(n+1)$.

Centered conditionalization and the above then entail that:

$$\mathrm{cr}_{t=n}(D) \quad = \quad \frac{N_{D_n} \cdot \mathrm{pr}_{D_n}}{N_{D_n} \cdot \mathrm{pr}_{D_n} + N_{S_n} \cdot \mathrm{pr}_{S_n}} \tag{9.39}$$

$$\begin{aligned} &= \frac{N_{D_n} \cdot \mathrm{pr}_{D_n}}{N_{D_n} \cdot \mathrm{pr}_{D_n} + N_{S_n} \cdot \mathrm{pr}_{D_n} \cdot (n+1)/k} \\ &= \frac{\mathrm{pr}_{D_n}}{\mathrm{pr}_{D_n} + \mathrm{pr}_{D_n} \cdot (n+1)/k} \\ &= \frac{1}{1+(n+1)/k}. \end{aligned}$$

Thus:

$$\lim_{n\to\infty} (\mathrm{cr}_{t=n}(D)) = \lim_{n\to\infty} \left(\frac{1}{1+\frac{n+1}{k}} \right) = 0. \tag{9.40}$$

9.16 TSC Satisfies the Probability Axioms

The probability axioms are:

1. $\forall a, p(a) \geq 0$

2. $p(\Omega) = 1$

3. $p(a \cup b) = p(a) + p(b),\ \text{if } a \cap b = \emptyset$

To see that TSC satisfies the probability axioms:

1. $\forall a, cr_e^{TSC}(a) \geq 0.$

Recall the formula for $cr_e^{TSC}(a)$:

$$cr_e^{TSC}(a) = \frac{\overset{\mathrm{def}}{\sum_{(i|c_i \in a)}} cr(\bar{c}_i) \cdot m_{\bar{c}_i}(c_i e|dts(cr))}{\overset{\mathrm{def}}{\sum_j} cr(\bar{c}_j) \cdot m_{\bar{c}_j}(c_j e|dts(cr))} \tag{9.41}$$

Now, $cr(\bar{c}_i) \geq 0$, and $m_{\bar{c}_i}(c_i e|dts) \geq 0$. Since the summation and multiplication of non-negative terms yields non-negative terms, $cr_e^{TSC}(a)$ is non-negative.

2. $cr_e^{TSC}(\Omega) = 1.$

$$\begin{aligned} cr_e^{TSC}(\Omega) &= \frac{\overset{\mathrm{def}}{\sum_i} cr(\bar{c}_i) \cdot m_{\bar{c}_i}(c_i e|dts(cr))}{\overset{\mathrm{def}}{\sum_j} cr(\bar{c}_j) \cdot m_{\bar{c}_j}(c_j e|dts(cr))} \\ &= 1. \end{aligned} \tag{9.42}$$

3. If $a \cap b = \emptyset$, then $cr_e^{TSC}(a \cup b) = cr_e^{TSC}(a) + cr_e^{TSC}(b)$.

$$\begin{aligned} cr_e^{TSC}(a \cup b) &\overset{\text{def}}{=} \sum_{(i|c_i \in a \cup b)} cr(\overline{c}_i) \cdot m_{\overline{c}_i}(c_i e|dts(cr)) \cdot N \\ &\overset{\text{def}}{=} \sum_{(i|c_i \in a)} cr(\overline{c}_i) \cdot m_{\overline{c}_i}(c_i e|dts(cr)) \cdot N \\ &\quad + \sum_{(i|c_i \in b)} cr(\overline{c}_i) \cdot m_{\overline{c}_i}(c_i e|dts(cr)) \cdot N \\ &= cr_e^{TSC}(a) + cr_e^{TSC}(b). \end{aligned} \tag{9.43}$$

9.17 TSC Reduces to Standard *De Dicto* Conditionalization

In standard contexts, a subject has one and only one temporal successor at each of their doxastic worlds, and the object and evidence are *de dicto* propositions. In these cases, TSC reduces to classical Bayesianism.

Let A and E be *de dicto* propositions. TSC assigns the following credences:

$$\begin{aligned} cr_E^{TSC}(A) &\overset{\text{def}}{=} \sum_{(i|c_i \in A)} cr(\overline{c}_i) \cdot m_{\overline{c}_i}(c_i E|dts(cr)) \cdot N \\ &\overset{\text{def}}{=} \sum_{(j|W_j \in A)} cr(W_j) \cdot \sum_{(l|c_k \in W_j)} m_{W_j}(c_k E|dts) \cdot N \end{aligned} \tag{9.44}$$

Now, we can ignore all j for which $cr(W_j) = 0$, since they make no contribution. The other W_j are doxastic worlds. Since by assumption we have one and only one temporal successor at each doxastic world, it follows that $m_{W_j}(dts) = 1$, and so $m_{W_j}(c_k E|dts) = m_{W_j}(c_k Edts)$. It further follows that $\sum_{(l|c_k \in W_j)} m_{W_j}(c_k Edts)$ will equal 1 if E contains W_j, since W_j is a doxastic world and we're guaranteed a single temporal successor there. On the other hand, $\sum_{(l|c_k \in W_j)} m_{W_j}(c_k Edts)$ will equal 0 if E doesn't contain W_j. So we can replace $\sum_{(l|c_k \in W_j)} m_{W_j}(c_k Edts)$ with

$cr(E|W_j)$, since it takes on the same values. This gives us:

$$\begin{aligned}\overset{\text{def}}{\sum_{(j|W_j\in A)}} cr(W_j)\cdot \sum_{(l|c_k\in W_j)} m_{W_j}(c_k E|dts)\cdot N &= \overset{\text{def}}{\sum_{(j|W_j\in A)}} cr(W_j)cr(E|W_j)\cdot N \\ &= \overset{\text{def}}{\sum_{(j|W_j\in A)}} cr(EW_j)\cdot N \qquad (9.45) \\ &= cr(EA)\cdot N.\end{aligned}$$

Now, the value of N is:

$$\begin{aligned} N &= \frac{1}{\overset{\text{def}}{\sum_i} cr(\bar{c}_i)\cdot m_{\bar{c}_i}(c_i e|dts)} \qquad (9.46) \\ &= \frac{1}{\overset{\text{def}}{\sum_j} cr(W_j)\cdot \sum_{(l|c_k\in W_j)} m_{W_j}(c_k E|dts)}\end{aligned}$$

Applying the above argument to this yields:

$$N = \frac{1}{E}. \qquad (9.47)$$

So:

$$\begin{aligned} cr_E^{TSC}(A) &= cr(EA)\cdot N \qquad (9.48) \\ &= \frac{cr(EA)}{cr(E)} \\ &= cr(A|E),\end{aligned}$$

which is the formula given by classical Bayesianism.

9.18 TSC Satisfies the Learning Principles

First we'll show that TSC satisfies LP$_1$. Then we'll see how to extend this result to LP$_n$.

9.18.1 TSC Satisfies LP_1

We want to show that TSC satisfies LP_1: A rational updating rule R must be such that if (i) all of a subject's doxastic alternatives have temporal successors, and (ii) none of these successors suffer from *de dicto* information loss, then her *de dicto* credences will lie in the span of the credences R assigns given her doxastic evidence."

Consider the probability functions p, and a series of probability functions p_i. As Van Fraassen (1995) notes, if we can always find coefficients α_i such that the following three conditions hold, then p is a mixture of p_is, and thus p lies in the span of the p_is.

(a)

$$\sum_i \alpha_i = 1 \tag{9.49}$$

(b)

$$\forall i(\alpha_i \in [0,1]) \tag{9.50}$$

(c)

$$p(\cdot) \overset{\text{def}}{=} \sum_i \alpha_i p_i(\cdot) \tag{9.51}$$

In the case of interest, p is the subject's current credence function cr, the p_is are the credence functions TSC assigns her given some doxastic evidence e_i, $cr_{e_i}^{TSC}$. If we can find α_is that satisfy (a), (b) and (c) when the prerequisites (i) and (ii) hold, then we've shown that TSC satisfies LP_1.

Define α_i as the inverse of the TSC normalization factor given evidence e_i, $N(e_i)$:

$$\alpha_i = \frac{1}{N(e_i)} \overset{\text{def}}{=} \sum_j cr(\overline{c}_j) \cdot m_{\overline{c}_j}(c_j e_i | dts). \tag{9.52}$$

This satisfies (a), (b) and (c).

(a).

$$\begin{aligned}\overset{\text{def}}{\sum_{(i|e_i \cup dts \neq \emptyset)}} \alpha_i &= \overset{\text{def}}{\sum_{(i|e_i \cup dts \neq \emptyset)}} \left(\overset{\text{def}}{\sum_j} cr(\overline{c}_j) \cdot m_{\overline{c}_j}(c_j e_i | dts) \right) \\ &= \overset{\text{def}}{\sum_j} cr(\overline{c}_j) \left(\overset{\text{def}}{\sum_{(i|e_i \cup dts \neq \emptyset)}} m_{\overline{c}_j}(c_j e_i | dts) \right)\end{aligned} \tag{9.53}$$

Since all of one's doxastic temporal successors will be compatible with some kind of evidence, the sum over i will get a non-zero contribution from $m_{\overline{c}_j}(c_j e_i|dts)$ for some e_i if $c_j \in dts$, but only from one e_i, since c_j can't belong two different e_is—evidence is mutually exclusive. So if $c_j \in dts$, the value of the sum over is will be $m_{\overline{c}_j}(c_j|dts)$. If $c_j \notin dts$, then the value of the sum over i will be 0. But if $c_j \notin dts$, then the value of $m_{\overline{c}_j}(c_j|dts)$ will be 0 as well, so we can just replace the sum over i with $m_{\overline{c}_j}(c_j|dts)$:

$$\overset{\text{def}}{\sum_j} cr(\overline{c}_j) \left(\overset{\text{def}}{\sum_{(i|e_i \cup dts \neq \emptyset)}} m_{\overline{c}_j}(c_j e_i | dts) \right) = \overset{\text{def}}{\sum_j} cr(\overline{c}_j) \cdot m_{\overline{c}_j}(c_j|dts) \tag{9.54}$$

Rearranging this sum over all centered worlds into a sum over all worlds and a sum over the centered worlds at each world, we get:

$$\begin{aligned}\overset{\text{def}}{\sum_j} cr(\overline{c}_j) \cdot m_{\overline{c}_j}(c_j|dts) &= \overset{\text{def}}{\sum_k} cr(W_k) \cdot \sum_{(l|c_l \in W_k)} m_{W_k}(c_l|dts) \\ &= \overset{\text{def}}{\sum_k} cr(W_k) \\ &= 1.\end{aligned} \tag{9.55}$$

(b). The α_is can't be less than 0, since they are formed from sums of products of positive terms. And they can't be greater than 1 either. $m_{\overline{c}_j}(c_j e_i|dts) \leq 1$, so:

$$\alpha_i = \overset{\text{def}}{\sum_j} cr(\overline{c}_j) \cdot m_{\overline{c}_j}(c_j e_i|dts) \leq \overset{\text{def}}{\sum_j} cr(\overline{c}_j) = 1. \tag{9.56}$$

(c).

$$
\begin{aligned}
\sum_{(i|e_i\cup dts\neq\emptyset)}^{\text{def}} \alpha_i \cdot cr_{e_i}^{TSC}(A) &= \sum_{(i|e_i\cup dts\neq\emptyset)}^{\text{def}} \alpha_i \sum_{(j|c_j\in A)}^{\text{def}} cr(\overline{c}_j)\cdot m_{\overline{c}_j}(c_je_i|dts)\cdot N(e_i)\\
&= \sum_{(i|e_i\cup dts\neq\emptyset)}^{\text{def}} \left(\sum_{(j|c_j\in A)}^{\text{def}} cr(\overline{c}_j)\cdot m_{\overline{c}_j}(c_je_i|dts)\right) \qquad (9.57)\\
&= \sum_{(i|e_i\cup dts\neq\emptyset)}^{\text{def}} \left(\sum_{(k|W_k\in A)}^{\text{def}} cr(W_k)\cdot \sum_{(l|c_l\in W_k)} m_{W_k}(c_le_i|dts)\right)\\
&= \sum_{(k|W_k\in A)}^{\text{def}} cr(W_k)\left(\sum_{(i|e_i\cup dts\neq\emptyset)}^{\text{def}} \sum_{(l|c_l\in W_k)} m_{W_k}(c_le_i|dts)\right)\\
&= \sum_{(k|W_k\in A)}^{\text{def}} cr(W_k)\left(\sum_{(l|c_l\in W_k)} m_{W_k}(c_l|dts)\right)\\
&= \sum_{(k|W_k\in A)}^{\text{def}} cr(W_k)\cdot m_{W_k}(dts|dts)\\
&= \sum_{(k|W_k\in A)}^{\text{def}} cr(W_k)\\
&= cr(A).
\end{aligned}
$$

So TSC satisfies LP$_1$.

9.18.2 TSC Satisfies LP$_n$

We've seen that TSC satisfies LP$_1$. Given this, it follows that that TSC satisfies LP$_n$ as well. To see why, first recall from chapter 6 why LP$_1$ is usually weaker than LP$_n$. LP$_1$ entails, given that the antecedent conditions are satisfied for each of the n steps, that a subject's current *de dicto* credences will lie in the span of the credences R assigns to her given any n-pieces of evidence formed the following way: the first piece of evidence is that of one of her doxastic temporal successors, the second piece of evidence is that of one of *their* doxastic temporal successors, and so on. Call the collection of such sequences of evidence her "doxastically-iterative n-step evidence". This is usually a weaker constraint than LP$_n$ imposes, since LP$_n$ requires the subject's credences to lie in the span of the credences R

assigns given her doxastic n-step evidence, which is a subset of her doxastically-iterative n-step evidence.

Now, a subject's doxastic n-step evidence is usually a subset of her doxastically-iterative n-step evidence because her successors can have doxastic alternatives which aren't doxastic successors of hers, and so her successors can think they might get evidence which she knows she'll never get. But in the case of TSC, this isn't a possibility, since the successors of a subject who updates using TSC will never have doxastic alternatives that aren't doxastic temporal successors of her predecessor: TSC will assign any centered worlds not in her predecessor's *dts* a credence of 0. So if the subject updates using TSC, however, her doxastic n-step temporal successors *will* be identical to her nth-iteration doxastic temporal successors. And so if TSC satisfies LP_1, it will satisfy LP_n as well.

So TSC satisfies both of the Learning Principles.

9.19 ESC Satisfies LP_1

Section 9.18 provides a proof that TSC satisfies LP_1. By replacing *dts* with *des* throughout, and replacing talk of temporal successors with talk of epistemic ones, that proof turns into a proof that ESC satisfies LP_1.

Bibliography

Albert, David. 2001. *Time and Chance.* Harvard University Press.

Arntzenius, Frank. 2002. "Reflections on Sleeping Beauty." *Analysis* 62:53–61.

Arntzenius, Frank. 2003*a*. "Self-locating Beliefs, Reflection, Conditionalization and Dutch Books." *Journal of Philosophy* 100:356–370.

Arntzenius, Frank. 2003*b*. "What Cloth?" Unpublished Manuscript.

Arntzenius, Frank and Ned Hall. 2003. "On What We Know About Chance." *British Journal for the Philosophy of Science* 54:171–179.

Bartha, Paul and Christopher Hitchcock. N.d. "No one knows the date or the hour: an unorthodox application of Rev. Bayes' Theorem." *Philosophy of Science (Proceedings).* Forthcoming.

Berndl, K, M Daumer, D Durr, S Goldstein and N Zanghi. 1995. "A Survey of Bohmian Mechanics." *Il Nuovo Cimento* 110:737–750.

Christensen, David. 1991. "Clever Bookies and Coherent Beliefs." *The Philosophical Review* 100:229–247.

Dorr, Cian. 2002. "Sleeping Beauty: in defense of Elga." *Analysis* 62:292–296.

Earman, John. 1986. *A Primer on Determinism.* Springer.

Earman, John. 1992. *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory.* MIT Press.

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

Elga, Adam. 2004. "Defeating Dr. Evil with self-locating belief." *Philosophy and Phenomenological Research* 69:383–396.

Hajek, Alan. 2003. "What Conditional Probabilities Could Not Be." *Synthese* 137:273–323.

Hall, Ned. 1994. "Correcting the Guide to Objective Chance." *Mind* 103:505–517.

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

Halpern, Joseph Y. 2001. Lexicographic Probability, Conditional Probability, and Nonstandard Probability. In *Proceedings of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge.* Morgan Kaufmann Publishers.

Halpern, Joseph Y. 2005. *Oxford Studies in Epistemology.* Vol. 1 Oxford University Press chapter Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems, pp. 111–142.

Halpern, Joseph Y and M R Tuttle. 1993. "Knowledge, probability and adversaries." *Journal of the ACM* 40:917–962.

Hawthorne, John. 2007. "Eavesdroppers and Epistemic Modals." Unpublished Manuscript.

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

Hoefer, Carl. 2005. "The Third Way on Objective Probability: A Skeptic's Guide to Objective Chance." Unpublished Manuscript.

Howson, Colin and Peter Urbach. 1993. *Scientific Reasoning: The Bayesian Approach.* 2nd ed. Open Court Publishing Company.

Jeffrey, Richard. 1983. *The Logic of Decision.* 2nd ed. University of Chicago Press.

Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *The Philosophical Review* 88:513–543.

Lewis, David. 1986*a*. *On The Plurality of Worlds.* Blackwell.

Lewis, David. 1986*b*. *Philosophical Papers.* Vol. 2 Oxford University Press chapter A Subjectivist's Guide to Objective Chance.

Lewis, David. 1994. "Humean Supervenience Debugged." *Mind* 103:473–490.

Lewis, David. 2001. "Sleeping Beauty: Reply to Elga." *Analysis* 61:171–176.

Lewis, David. 2004. "How Many LIves Has Schrodingers Cat?" *Australasian Journal of Philosophy* 82:3–22.

Loewer, Barry. 2001. "Determinism and Chance." *Studies in the History of Modern Physics* 32:609–620.

Maher, Patrick. 2006. "The Concept of Inductive Probability." *Erkenntnis* 65:185–206.

Meacham, Christopher J G. 2005. "Three Proposals Regarding a Theory of Chance." *Philosophical Perspectives* 19:281–307.

Meacham, Christopher J G. 2006. “Sleeping Beauty and the Dynamics of *De Se* Beliefs.” *Philosophical Studies* .

Miller, David. 1966. “A Paradox of Information.” *British Journal for the Philosophy of Science* 17:59–61.

Norton, John D. 2003. “Causation as Folk Science.” *Philosophers Imprint* 3:1–22.

Price, Huw. 1996. *Time’s Arrow and Archimedes’ Point: New Directions for the Physics of Time.* Oxford University Press.

Strevens, Michael. 1995. “A Closer Look at the New Principle.” *British Journal for the Philosophy of Science* 46:545–561.

Strevens, Michael. 2004. “Bayesian Confirmation Theory: Inductive Logic, or Mere Inductive Framework?” *Synthese* 141:365–379.

Thau, Michael. 1994. “Undermining and Admissibility.” *Mind* 103:491–503.

Tolman, R C. 1979. *The Principles of Statistical Mechanics.* Dover Publications.

Tumulka, R and N Zanghi. 2005. “Thermal Equilibrium Distribution of Wavefunctions.” arXiv:quant-ph/0309021v2.

Van Fraassen, Bas. 1995. “Belief and the Problem of Ulysses and the Sirens.” *Philosophical Studies* 77:7–37.

Vranas, Peter. 2002. “Whos Afraid of Undermining? Why the Principal Principle might not contradict Humean Supervenience.” *Erkenntnis* 57:151–174.

Vranas, Peter. 2004. “Have your cake and eat it too: The Old Principal Principle reconciled with the New.” *Philosophy and Phenomenological Research* 69:368–382.

Wallace, David. 2006. “Epistemology Quantized: circumstances in which we should come to believe in the Everett interpretation.” *British Journal for the Philosophy of Science* 57:655–689.

Weatherson, Brian. 2005. “Should we Respond to Evil with Indifference?” *Philosophy and Phenomenological Research* 70:613–635.

Weisberg, Jonathan. 2006. Credence and Credibility: the Commonsense Approach to Probabilistic Epistemology PhD thesis Rutgers University.

Weisberg, Jonathan. 2007*a*. “Conditionalization, Reflection, and Self-Knowledge.” *Philosophical Studies* .

Weisberg, Jonathan. 2007*b*. “Conditionalization without Reflection.” Unpublished Manuscript.

Wilson, Mark. 1993. “There’s a Hole and a Bucket, Dear Leibniz.” *Midwest Studies in Philosophy* 18:202–241.

Xia, Z. 1992. “The Existence of Noncollision Singularities in the N-body Problem.” *Annals of Mathematics* 135:411–468.

Vita

Christopher J. G. Meacham

2007	Ph.D. in Philosophy, Rutgers University
1999	B.A. in Physics and Philosophy, Reed College

2007-	Assistant Professor, Department of Philosophy, University of Massachusetts, Amherst
2006-2007	Bersoff Faculty Fellow, Department of Philosophy, NYU
2003, 2005	Teaching Assistant, Department of Philosophy, Rutgers University