1. Putting a Value on Beauty

Rachael Briggs

1. STAGE SETTING

Adam Elga's Sleeping Beauty problem is an instance of a more general puzzle: what is the relevance of purely *de se* information to *de dicto* beliefs? In part 1, I explain the distinction between *de dicto* and *de se* beliefs (section 1.1), remind readers of the standard Bayesian story about belief updating (section 1.2), and summarize the debate between halfers and thirders about how to extend the standard Bayesian story to the Sleeping Beauty case (section 1.3).

In part 2, I explain why the trouble posed by the Sleeping Beauty problem—and by *de se* information in general—is much deeper than it first appears. I build a framework that gets around the trouble by dividing agents' belief functions into a *de dicto* component and a *de se* component. Using my framework, I formulate a 'Halfer Rule' and a 'Thirder Rule'—extensions of the halfer and thirder solutions to the Sleeping Beauty problem.

The Halfer Rule and the Thirder Rule both seem appealing, but unfortunately, they are incompatible. Which is preferable? I consider two ways of resolving the question: Dutch books in part 3, and scoring rules in part 4. I argue that on both ways of resolving the conflict, whether one should prefer the Halfer Rule or the Thirder Rule depends on one's views about decision theory. Evidential decision theorists should prefer the Halfer Rule, while causal decision theorists should prefer the Thirder Rule.

In part 5, I argue that there is a consideration that favors the Thirder Rule over the Halfer Rule independently of any scoring considerations. The Thirder Rule is stable in a way that the Halfer Rule is not. Enriching an agent's belief worlds with irrelevant information will dramatically change the advice of the Halfer Rule. Since the Thirder Rule is closely connected to causal decision theory and the Halfer Rule is closely connected to evidential decision theory, the stability of the Thirder Rule constitutes a reason for preferring evidential decision theory to causal decision theory.

1.1. *The* de dicto–de se *Distinction*

In the terminology of Lewis (1979), *de dicto* beliefs concern only what the world is like. A person's *de dicto* beliefs might include the belief that the meek will inherit the Earth, the belief that flossing prevents tooth decay, or

I would like to thank Ned Hall, Alan Hajek, Vann McGee, and Bob Stalnaker for their extensive comments on earlier drafts of this essay.

the belief that a Republican won the 1992 US presidential election.[1] If two inhabitants of the same possible world have the same *de dicto* beliefs, they are either both right or both wrong. Furthermore, if a person has the same *de dicto* beliefs at different times in the same world, then she is either right both times or wrong both times:

De se beliefs, by contrast, concern the believer's location (in addition to what the world is like).[2] A person's *de se* beliefs might include the belief that today is Tuesday, the belief that her shopping cart contains a leaky bag of sugar (Perry, 1971), or the belief that she will be devoured by ravenous dogs (Quine, 1969). It's possible for two inhabitants of the same world to have the same *de se* beliefs, but for one to be right and the other to be wrong; Napoleon might truly believe, and I might falsely believe, the *de se* content expressed by the sentence "I am Napoleon." Furthermore, it's possible for an inhabitant of a world to have the same *de se* beliefs at different times in the same world, and for her to be right at some times and wrong at others; I might truly believe an Monday, and falsely believe on Tuesday, the *de se* content expressed by the sentence "It is Monday." A *de dicto* belief can be represented using the proposition that its content is true—i.e. the set of possible worlds where its content is true. A *de se* belief cannot.

Why not? Consider the example of Rudolph Lingens, an amnesic lost in the Main Library at Stanford (Perry, 1974). Lingens happens upon a biography of himself and discovers all sorts of interesting *de dicto* facts about Lingens. We might suppose (embellishing Perry's story) that the book was written by an oracle and contains the sentence "Lingens is in aisle five, floor six, Main Library, Stanford, on October 1, 1977." But no matter how much *de dicto* information Lingens acquires by reading the book, he is missing the crucial *de se* information that would allow him to find his way out of the library. Until he learns that *he* is Rudolph Lingens, and that *today* is October 1, 1977, he is lost.

[1] Whether these beliefs are truly *de dicto* is open to debate. Perhaps 'the Earth' really means something like 'the Earth-like planet inhabited by *me*'. If so, then the belief that the meek will inherit the Earth is irreducibly *de se*. In a universe with two very similar planets, both called Earth, I might believe truly that the meek will inherit the Earth (because the meek will inherit the Earth-like planet that I inhabit), while my twin on the other planet might believe falsely that the the meek will inherit the Earth (because the meek will not inherit the Earth-like planet that she inhabits). It is hard to formulate an English sentence that uncontroversially expresses a *de dicto* claim. But if there are no purely *de dicto* beliefs, then this is all the more reason to attend to the epistemic role of *de se* evidence.

[2] Whether these beliefs are truly *de se* is open to debate. Millikan (1990) claims that there are no self-locating contents, only different roles that a belief might play in the believer's mental life. She objects to one reason for positing self-locating beliefs—the theory that an agent's self-locating beliefs explain her ability to act on her other beliefs. Millikan's argument is persuasive, but it does not establish that there are no self-locating beliefs. Rather, it establishes that self-locating beliefs are not sufficient to explain an agent's ability to act on her beliefs. It may be that irreducibly *de se* beliefs explain some (but not all) of the behavior that *de dicto* beliefs cannot account for alone.

The contents of *de se* beliefs can be represented using *centered propositions*, or sets of centered worlds—where a centered world is an ordered pair consisting of a world *W* and a *center*, or the spatiotemporal location of an individual, in *W*.[3] (In the remainder of the paper, I will refer to the worlds and propositions used to represent *de dicto* belief as *uncentered*. Every uncentered proposition *A* is equivalent to some centered proposition—roughly, the set of centers located in worlds where *A* is true—but not every centered proposition is equivalent to an uncentered proposition.)

We can use the terms '*de dicto*' and '*de se*' to describe properties of overall belief states, as well as properties of individual beliefs. Say that a person suffers from irreducibly *de se* ignorance just in case some of her doxastically possible worlds contain more than one doxastically possible center. When someone suffers from irreducibly *de se* ignorance, uncentered worlds are too coarse-grained to distinguish among her doxastic alternatives. Her epistemic state is best represented using centered worlds. Say that someone's ignorance is purely *de dicto* just in case she is ignorant, but her ignorance is not irreducibly *de se*.

One might read Perry's Lingens story as an example of someone suffering irreducibly *de se* ignorance. For all Lingens knows, there are two people lost in the library who satisfy all of his *de se* beliefs. For all he knows, these people are facing indistinguishable shelves of books, thinking indistinguishable thoughts, and wearing shoes of indistinguishable colors. Described this way, Lingens's ignorance is irreducibly *de se*: some of his doxastically possible worlds contain two doxastically possible centers.

Note that not all ignorance about one's identity or location is irreducibly *de se*. In order for Lingens's ignorance to be irreducibly *de se*, it is not sufficient for him to be unsure whether he is Rudolf Lingens or Bernard J. Ortcutt. If Lingens is unsure whether he is Rudolf Lingens or Bernard J. Ortcutt, believes that he is lost in a library, and believes that exactly one person is lost in a library, then his ignorance can be characterized in purely *de dicto* terms. He has some doxastic alternatives where Lingens is lost in the library (and Ortcutt is not), and some where Ortcutt is lost in the library (and Lingens is not). These alternatives can easily be represented by uncentered worlds.

[3] Are spatiotemporal locations really enough to capture the intuitive concept of a center? To pre-empt any mischief on the part of my readers, I'll go ahead and formulate the counterexample from hell: suppose some sorcerer turns me into an extended ghost that can superimpose itself on ordinary matter. The sorcerer then packs me into her convenient time machine and sends me back to my own past. There, I superimpose my ghostly body on my older, more substantial body. My 'later', ghostly self experiences different thoughts and qualia than my 'earlier', physical self. There seem to be two centers here, although there is only one individual (me) who occupies one spatiotemporal location. One might refine my rough-and-ready definition either by adding in a clause about the unity of qualia, or by quibbling about personal timelines in the definition of 'spatiotemporal location'. For the purposes of this essay, characterizing centers as spatiotemporal locations will work well enough.

1.2. De dicto *Confirmation Theory*

Standard Bayesian confirmation theory is built to cope with *de dicto* ignorance rather than irreducibly *de se* ignorance. It posits credences that attach to uncentered propositions. Updating is bound by the following norm (where Cr is the agent's original credence function, E is her new evidence, Cr_E is her credence function updated on E, and A is any proposition):

$$\text{Conditionalization} : Cr_E(A) = Cr(A|E) = \frac{Cr(A \,\&E)}{Cr(E)}$$

A and E are assumed to be uncentered propositions.

Conditionalization is a useful rule in most situations—but not in all. Sometimes conditionalization is too restrictive, as in cases of memory loss. If an agent forgets a proposition A, she cannot possibly update by conditionalization. Her old credence in A conditional on absolutely any proposition was 1, but if she forgets whether A, then her new credence in A is less than one.

At other times, conditionalization is not restrictive enough, as in cases where the agent learns something incompatible with her old beliefs. Where $Cr(E)$ is 0, the ratio $\frac{Cr(A\,\&\,E)}{Cr(E)}$ is undefined. Mathematically speaking, it's easy to pick out a conditional credence function $Cr(\cdot|E)$ which is equal to $\frac{Cr(A\,\&\,E)}{Cr(E)}$ where $E > 0$ and well defined even where $Cr(E) = 0$ (see McGee 1994; Hájek 2003). The trouble is that there are *many* possible ways to do so. If $Cr(E) = 0$, then as long as A doesn't entail E, the agent's real-valued credences allow Cr $(A|E)$ to be anything. Unless we have some ways of narrowing down the range of acceptable conditional credence functions, conditionalization doesn't amount to a real constraint.

One more note on the decision-theoretic apparatus: throughout this essay, I will assume that the domains of agents' credence functions can be represented using finitely many possible worlds. I'll be treating possible worlds not as maximally specific (descriptions of) states of affairs, but as (descriptions of) states of affairs that are sufficiently rich for the theoretical purposes at hand.

1.3. *Sleeping Beauty*

Elga (2000) has argued that irreducibly *de se* ignorance generates exceptions to *de dicto* conditionalization. He supports this claim with the following *Sleeping Beauty* example (2000: 143), addressing the protagonist, whom I will call 'Beauty', in the second person:

> Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking.
> When you are first awakened, to what degree ought you believe that the outcome of the coin toss is heads?

The first waking will occur on Monday, and the second, if it occurs at all, will occur on Tuesday. We can assume that the coin flip occurs after the Monday waking and before the Tuesday one. When Beauty wakes, one of her doxastically possible worlds (the one where the coins lands tails) will contain more than one doxastically possible center (the Monday center and the Tuesday center). Therefore, upon waking, Beauty will suffer from irreducibly *de se* ignorance, and her credence function upon waking is best represented as a probability function over a set of centered propositions.

Elga claims that Beauty should place credence 1/3 in the proposition that the coin lands heads. More specifically, her credence function on waking up (which I will call Cr^{up}) should assign the following values to centered propositions:

	Heads	**Tails**
Monday	1/3	1/3
Tuesday	0	1/3

Each column in the table represents an uncentered proposition, while each entry represents a centered proposition. I will follow Elga's scheme for naming these centered propositions: H_1 = Monday, heads; T_1 = Monday, tails; and T_1 = Tuesday, tails. *Heads* is the uncentered proposition that the coin lands heads; *Tails* is the uncentered proposition that the coin lands tails. Likewise, *Monday* is the centered proposition that it is Monday; *Tuesday* is the centered proposition that it is Tuesday. Thus, according to Elga $Cr^{up}(H_1) = Cr^{up}(T_1) = Cr^{up}(T_2) = 1/3$. Call the adoption of this credence function *thirding* and those who support it *thirders*.

Elga's argument for thirding is as follows. First, he claims that

$$Cr^{up}(T_1) = Cr^{up}(T_2) \tag{1}$$

1 is a consequence of the following highly restricted indifference principle:
Weak Indifference If two centered worlds W_1 and W_2 are subjectively indistinguishable, and the same propositions are true in both, then W_1 and W_2 should receive equal credence.

Next, let Cr^+ be the credence function that it would be rational for Beauty to adopt after waking up and coming to believe *Monday*. Then

$$Cr^+(H_1) = Cr^+(T_1) = 1/2 \tag{2}$$

Why so? We can imagine that Beauty knows the coin will not be flipped until after she forms Cr^+. When she forms Cr^+, she learns that the coin (which is fair), has not yet been flipped, so she should assign equal credence to H_1 and T_1. Since Beauty should conditionalize on her total evidence, says Elga,

$$Cr^{up}(H_1|Monday) = Cr^{up}(T_1|Monday) = 1/2 \tag{3}$$

Together, 1 and 2 fix the probabilities of all the cells in the table.

If Elga is right, then the *de dicto* version of conditionalization is wrong. According to Elga, Beauty's Sunday credence in *Heads* should be 1/2, while her Monday credence in *Heads* should be 1/3. But Beauty gains no uncentered evidence between Sunday and Monday. According to conditionalization, her Monday credence in *Heads* should be 1/2 rather than 1/3.

Contra Elga, Lewis (2001) claims that the correct value of Cr^{up} is:

	Heads	**Tails**
Monday	1/2	1/4
Tuesday	0	1/4

Call the adoption of this credence function *halfing* and those who support it *halfers*.

Lewis argues for halfing on the grounds that Beauty acquires no new information between Sunday and Monday. Arntzenius (2002) points out that Lewis should invoke a second premise—that Beauty loses no information between Sunday and Monday. A halfer might support both premises by pointing out that Beauty neither loses nor gains *de dicto* information between Sunday and Monday. If any change in her *de se* evidence is irrelevant to her beliefs about the coin toss, then she ought to half.

Oddly enough, Lewis grants that *de se* evidence is sometimes relevant to *de dicto* beliefs. He claims that

$$Cr^{up}(H_1|Monday) = 2/3 \tag{4}$$

Like Elga, Lewis believes that Beauty should conditionalize on her total evidence. So once she learns that it's Monday,

$$Cr^{+}(H_1) = 2/3 \tag{5}$$

As we will see in part 2, halfers need not agree to 5, or grant that *de se* evidence is ever relevant to *de dicto* beliefs. In fact, I will argue in section 2.1 that they should do neither of these things.

2. RULES FOR HALFERS AND THIRDERS

Elga and Lewis disagree about how Beauty should update her beliefs when she receives new, irreducibly *de se* evidence, but they agree that she should conditionalize on her total evidence. Bostrom (2007) and Meacham (2008) have pointed out that this shared assumption is misguided.

I argued in section 1.2 that conditionalization is useful only when the agent's new evidence is compatible with her old beliefs. But even routine *de se* updating involves new *de se* evidence that is incompatible with the agent's old *de se* beliefs. To borrow an example from Meacham (2008), suppose I start out believing that it is noon, then glance at my clock and realize that it is 12:01. My new belief that it is 12:01 is incompatible with my old belief that it is noon.

Several authors suggest sophisticated ways around this problem. Meacham (2008) uses a 'hypothetical prior', which assigns probabilities to centered worlds outside the space of the agent's doxastic alternatives. Titelbaum (2008) proposes a system where sentences, rather than propositions, receive credences. I suggest a simpler solution: we can just divide the agent's belief state into one part that represents her uncentered information, and a second part that represents her additional centered information.

The first, uncentered part of the agent's belief state is a credence function Cr_u that ranges over uncentered propositions. We can think of Cr_u as the credence function the agent would endorse as the correct 'view from nowhere', given her priors and her evidence. The second, centered part of the agent's belief state is a function that takes each doxastically possible uncentered world W to a natural number N_W equal to the number of doxastically possible centers in W.

Like the agent's belief state, any new evidence that the agent acquires can be divided into centered and uncentered parts. (Throughout this chapter, I'll assume that evidence comes in the form of a centered proposition the agent learns with certainty.) The uncentered part of a centered evidence proposition E is the strongest uncentered proposition entailed by E—the set of all and only the uncentered worlds that contain centers in E. (I'll call this proposition $u(E)$.) Cr_u can then be updated by conditionalizing on $u(E)$ when the agent learns E. The centered part of E is a new function taking each world W to a natural number N'_W—the number of centers in W after the agent has updated.

Let $Cr_{@}$ be the agent's actual credence function. What is the appropriate relationship between Cr_u and $Cr_{@}$? Two answers suggest themselves.

2.1. *The Halfer Rule*

We might claim $Cr_{@}$ should simply coincide with Cr_u for all uncentered propositions. In other words, where A is any uncentered proposition, we might endorse the

$$\textbf{Halfer Rule}: Cr_{@}(A) = Cr_u(A)$$

The Halfer Rule has been endorsed (albeit in different notation) by Halpern (2005) and Meacham (2008).

As its name suggests, the Halfer Rule entails that Beauty should half. But unlike Lewis's version of halfing, the Halfer rule is incompatible with Lewis's equation 5. Lewis espoused 5 because he believed that Beauty should update by conditionalizing on her total evidence. But as we've seen, conditionalization is useless for *de se* updating. The Halfer Rule recommends conditionalizing the *uncentered portion* of one's credence function on the *uncentered portion* of one's total evidence, and then within each world, dividing one's credence among the doxastically possible centers. Since Beauty gains no relevant uncentered evidence when she learns that it's Monday, the Halfer Rule entails Elga's equation 2.

It's just as well that the Halfer Rule leads us to reject 5. Draper and Pust (2008) formulate a Dutch book against agents who satisfy 5. (For reasons that will become clear in section 3.2, their Dutch book does *not* automatically carry over to agents who satisfy 4.)

2.2. *The Thirder Rule*

We might embrace the following alternative to the Halfer Rule, where A is any uncentered proposition, W and W^* are variables ranging over doxastically possible uncentered worlds, and N_W is the number of centers in W:

$$\textbf{Thirder Rule}: Cr_{@}(A) = \frac{\sum_{W \in A} Cr_u(W)N_W}{\sum_{W^*} Cr_u(W^*)N_{W^*}}$$

Elga (2007) endorses the Thirder Rule (albeit in different notation).

You might think of the Thirder Rule (a bit fancifully) as recommending that you follow this procedure: first, use the *de dicto* portion of your evidence to calculate Cr_u, the credence which it would be appropriate to place in each doxastically possible uncentered world W if you occupied the 'view from nowhere'. Next, use the *de se* portion of your evidence to increase your credence in worlds with multiple centers. If a world contains n centers, then n-tuple your credence in it. Finally, re-normalize your credences so that they sum to 1.

Of course, the Thirder Rule doesn't *really* recommend following the above procedure. It recommends adopting the credences you would arrive at if you followed the procedure, but it doesn't say how you should arrive at them. As its name suggests, the Thirder Rule entails that Beauty should third.

Notice that although the Thirder Rule disagrees with the Halfer Rule about what Beauty should do upon waking up, it agrees with the Halfer Rule about what she should do upon waking up and learning that it's Monday. Once Beauty learns that it's Monday, she has only one doxastically possible center per doxastically possible uncentered world. Like the Halfer Rule, the Thirder Rule entails Elga's equation 2.

2.3. *Weak Indifference*

Neither the Halfer Rule nor the Thirder Rule tells agents how to divide their credences among multiple centers within uncentered worlds. Halfers and thirders might enrich their accounts by espousing Elga's Weak Indifference principle. Weak Indifference seems to capture correctly an important symmetry in cases of *de se ignorance*: when two centered worlds are subjectively indistinguishable, and when the same propositions are true in both, it's hard to see any reason for granting one greater credence than the other.

Elga (2004) proposes a version of the following strange example. I will assume that you, the reader, are not a brain in a vat. But suppose you come to believe (with certainty) that some scientists have created a brain in a vat which is a subjective duplicate of you. Weak Indifference mandates that you place credence 1/2 in the proposition that you are a brain in a vat. To generalize the point, conditional on there being a brain in a vat whose state is subjectively indistinguishable from your own, you should place credence 1/2 on the proposition that you are a brain in a vat.

If Elga is right, then Weak Indifference has practical as well as theoretical consequences (at least for people who believe themselves to inhabit strange brain-in-vat worlds). If you satisfy Weak Indifference, and you are convinced that the world contains a brain in a vat whose state is subjectively indistinguishable from your own, then you should exert considerable effort to prevent anyone from torturing that brain. For (assuming you are certain the world contains the brain in the vat, and no other subjective duplicates of you) you must place credence 1/2 in the proposition that it is you who will suffer if the brain is tortured. Elga accepts these consequences of Weak Indifference, but many would find them counterintuitive.

Weatherson (2005) argues that in addition to having counterintuitive consequences, Weak Indifference is insufficiently motivated. Symmetry considerations do not succeed in establishing Weak Indifference—to think that they do is to conflate risk with uncertainty. A risky proposition (for a particular agent) is one whose truth value she does not know, and to which she can reasonably assign a precise credence. An uncertain proposition (again, for a particular agent) is one whose truth value she does not know, and to which she cannot reasonably assign a precise credence. Weak Indifference entails that if an agent has two doxastically possible centered worlds in which the same uncentered propositions are true, then she should assign each of them some precise credence. But in at least some cases of irreducibly *de se* ignorance, Weatherson contends, the agent should assign some centered worlds no precise credence at all—she should treat her location in the world as a matter of uncertainty, rather than a matter of risk.

Finally, Weatherson (2005) points out that when an agent has doxastically possible worlds containing a countable infinity of subjective duplicates, Weak Indifference is incompatible with countable additivity. There is simply no additive way to distribute credence evenly over countably many individuals.

Halfers and thirders who find Weak Indifference intuitively appealing, but are worried by these problems, might get around them by adopting a weaker version of Weak Indifference.

Weaker Indifference If two centered worlds W_1 and W_2 are subjectively indistinguishable, the same propositions are true in both, *and W_1 and W_2 both receive precise nonzero credence*, then W_1 and W_2 should receive equal credence.

Weaker Indifference lacks the counterintuitive consequences of Weak Indifference, and is better supported by symmetry considerations. In Elga's example, when you become certain that the world contains a brain in a vat which is a subjective duplicate of you, Weaker Indifference permits you to place credence 1/2 in the proposition that you are a brain in a vat. But unlike Weak Indifference, it also permits you to remain certain that you are not a brain in a vat (or even to refuse to place any precise credence in the proposition that you are a brain in a vat). In Weatherson's examples, when you become certain that the world contains countably many individuals whose states are subjectively indistinguishable from your own, Weaker Indifference does not require you to distribute your credence evenly over all those individuals. It leaves you the options of distributing your credence evenly over a finite number of them, and of refusing to distribute your credence over them at all. And Weaker Indifference is neutral on the question of when you should treat your location as a matter of risk rather than uncertainty—it simply tells you what to do if you treat your location as a matter of risk.

3. DUTCH BOOKS

Two authors (Hitchcock, 2004; Halpern, 2005), suggest that we adjudicate between the Halfer Rule and the Thirder Rule using Dutch books. Oddly enough, they come up with diametrically opposed results. Hitchcock claims that Dutch book considerations favor a thirder approach, while Halpern claims that they favor a halfer approach.

Who's right? The answer depends on which decision theory we adopt. I'll argue that causal decision theorists should be thirders, while evidential decision theorists should be halfers. I'll begin in section 3.1 by articulating an important and sometimes overlooked constraint on what counts as a Dutch book, and illustrating the importance of the constraint by showing that Halpern's (2005) putative Dutch book violates it.

In section 3.2, I'll explain Hitchcock's (2004) Dutch book against halfers. I'll argue that Hitchcock's Dutch book only works if Beauty is a causal decision theorist. In section 3.3, I argue that if Beauty is an evidential decision theorist who thirds, she is vulnerable to a Dutch book (though not the one suggested by Halpern). At this point, I will have given Dutch book arguments against halfing (if you're a causal decision theorist) and thirding (if you're an evidential decision theorist).

In the remainder of part 3, I will prove a pair of converse Dutch book results: causal decision theorists who obey the Thirder Rule and evidential decision theorists who obey the Halfer Rule are immune to Dutch books. In section 3.4, I'll establish a claim that serves as an important part of both converse Dutch books: anyone who bets at what I call *thirder odds* is immune to a Dutch book. In section 3.5, I'll argue that causal decision theorists who obey the Thirder Rule, and evidential decision theorists who obey the Halfer Rule, should bet at thirder odds. Putting these two results together yields the conclusion that causal decision theorists who obey the Thirder Rule, and evidential decision theorists who obey the Halfer Rule, are immune to Dutch books.

3.1. *A Constraint on Dutch Books*

A Dutch book is a set of bets that an agent is willing to accept individually (where an agent's willingness to accept a bet cashed out in terms of the bet's having positive expected value), but that jointly result in a sure loss. If a set of bets is to count as a Dutch book, however, the agent must be willing to accept the bets even when she is fully informed about their nature. For instance, if you believe to degree 1/2 that it will rain tomorrow, it is no objection to your credence function that a bookie could exploit you by selling you the following 'bet':

a ticket that costs \$5 and pays \$10 in the case of rain	if it will not rain	(6)
a ticket that costs \$5 and pays \$10 in the case of no rain	if it will rain	

Bet 6 isn't a Dutch book because your willingness to accept it depends on your ignorance of the betting conditions. If you knew that you would be offered the no-rain side of the bet in case of rain, and the rain side in case of no rain, you wouldn't make the bet. It's not your incoherence that bet 6 is punishing, but your ignorance.

Authors sometimes phrase the no-deception requirement in terms of whether the bookie knows anything that the buyer does not, but this is misleading. The bookie's epistemic state is beside the point—after all, a bookie could deceive an agent unwittingly. What matters is whether the bookie's behavior matches up with the agent's expectations.

In cases of irreducibly *de se* ignorance, it's easy to run afoul of the no-deception requirement. Halpern (2005) claims that if Beauty thirds, she is vulnerable to Lewis's 1999 Dutch book against agents who violate *de dicto* conditionalization. On closer examination, Halpern's extension of Lewis's Dutch book turns out to require deception.

On Sunday, a bookie might sell Beauty the following bet:

$$\begin{array}{ll} \textit{Heads} & \$(15+\epsilon) \\ \textit{Tails} & \$(-15+\epsilon) \end{array} \tag{7}$$

When she wakes up on Monday, the bookie might sell her

$$\begin{array}{ll} \textit{Heads} & \$(-20+\epsilon) \\ \textit{Tails} & \$(10+\epsilon) \end{array} \tag{8}$$

Together, bets 7 and 8 seem to constitute a Dutch book. If the coin lands heads, Beauty wins $\$(15+\epsilon)$ on bet 7, but loses $\$(20-\epsilon)$ on bet 8, and if the coin lands tails, she loses $\$(15-\epsilon)$ on bet 7 and wins back only $\$(10+\epsilon)$ on bet 8. Either way, she loses a total of $\$(5-2\epsilon)$.

But Hitchcock (2004) points out that this setup isn't a Dutch book at all. What if the coin lands tails? Does the bookie offer bet 8 only once, on Monday, or does he offer it twice, once on Monday and once on Tuesday? If the bookie offers bet 8 only on Monday, then Beauty should refuse to bet. Once she learns that the bookie has approached her, she'll know that it's Monday, and her credence in *Heads* will be 1/2—meaning that bet 8 has negative expected value. On the other hand, if the bookie offers bet 8 on Tuesday as well, then Beauty will end up $\$(5+2\epsilon)$ richer when the coin lands tails.

The only way the bookie can trap Beauty into a sure loss is by deceiving her. If Beauty is falsely convinced that she'll be offered bet 8 on Tuesday (if she is awake), then she's guaranteed to lose money. But since this scenario involves deception about the nature of the bets, it's not a Dutch book.

3.2. *Hitchcock's Dutch Book*

Hitchcock (2004) suggests a Dutch book against halfers that seems to avoid the deception pitfall described above. After phrasing the no-deception requirement in terms of the bookie's epistemic state, he writes:

> There is one way in which the bookie can ensure that he has no information that is unavailable to Beauty: he can sleep with her. That is, he can place his first bet, go into a deep sleep when Beauty does, arrange to have himself awakened under the same protocol as Beauty, and sell a follow-up bet to Beauty whenever they wake up together. The bookie, like Beauty, will awaken having no idea whether it is the first or second awakening, having no idea whether an initial follow-up bet has already been placed. Thus he must sell the same bet to Beauty whenever they both wake up.

On Hitchcock's way of individuating bets, the same bet is offered at the same stakes in all subjectively indistinguishable centers. We can consider this a stipulation about how to use the word 'bet'. (It's not the only acceptable way to individuate bets, but it's an acceptable way. The important thing is that Beauty know what the setup is.)

Hitchcock then argues that Beauty is vulnerable to the following Dutch book if she halfs. On Sunday, the bookie offers Beauty:

$$\begin{array}{ll} \textit{Heads} & \$(-15+2\epsilon) \\ \textit{Tails} & \$(15+\epsilon) \end{array} \tag{9}$$

On Monday (and on Tuesday, if she wakes up), he offers her:

$$\begin{array}{ll} \textit{Heads} & \$(10+\epsilon) \\ \textit{Tails} & \$(-10+\epsilon) \end{array} \tag{10}$$

Together, bets 9 and 10 result in a sure loss of $\$(5-3\epsilon)$ for Beauty. If the coin lands heads, she loses $\$(15-2\epsilon)$ on bet 9, and wins back only $\$(10+\epsilon)$ on bet 10, which is made only once (on Monday). If the coin lands tails, she wins $\$(15+\epsilon)$ on bet 9, but loses $\$(20-2\epsilon)$ on bet 10, which is made twice (on Monday and on Tuesday).

But will Beauty really accept bet 10? Adopting a thirder's credences doesn't necessarily translate to betting at a thirder's odds. Arntzenius (2002), who anticipates a version of Hitchcock's Dutch book, points out that that Beauty will only accept bets like 10 if she is a causal decision theorist as well as a halfer. If Beauty is an evidential decision theorist, she won't think bet 10 is worth accepting.

Arntzenius supports his point by drawing an analogy between *Sleeping Beauty* and a two-person prisoner's dilemma game. But no such analogy is necessary; we can see that Arntzenius is right by carefully going through the expected utility calculations. Suppose that Beauty has just awoken, that she is unsure whether it's Monday or Tuesday, and that she is deciding whether to accept bet 10. (Call the centered proposition that she accepts bet 10 right now *Accept*, and call the centered proposition that she accepts the same bet on a different day *D*.) The evidential expected values of *Accept* and ¬ *Accept* are

$$\begin{aligned} V_E(\textit{Accept}) \quad &= Cr(\textit{Heads} \,\&\, D|\textit{Accept})V(\textit{Heads} \,\&\, D \,\&\, \textit{Accept}) \\ &\quad + Cr(\textit{Heads} \,\&\, \neg D|\textit{Accept})V(\textit{Heads} \,\&\, \neg D \,\&\, \textit{Accept}) \\ &\quad + Cr(\textit{Tails} \,\&\, D|\textit{Accept})V(\textit{Tails} \,\&\, D \,\&\, \textit{Accept}) \\ &\quad + Cr(\textit{Tails} \,\&\, \neg D|\textit{Accept})V(\textit{Tails} \,\&\, \neg D \,\&\, \textit{Accept}) \end{aligned}$$

$$\begin{aligned} V_E(\neg \textit{Accept}) \quad &= Cr(\textit{Heads} \,\&\, D|\neg \textit{Accept})V(\textit{Heads} \,\&\, D \,\&\, \neg \textit{Accept}) \\ &\quad + Cr(\textit{Heads} \,\&\, \neg D|\neg \textit{Accept})V(\textit{Heads} \,\&\, \neg D \,\&\, \neg \textit{Accept}) \\ &\quad + Cr(\textit{Tails} \,\&\, D|\neg \textit{Accept})V(\textit{Tails} \,\&\, D \,\&\, \neg \textit{Accept}) \\ &\quad + Cr(\textit{Tails} \,\&\, \neg D|\neg \textit{Accept})V(\textit{Tails} \,\&\, \neg D \,\&\, \neg \textit{Accept}) \end{aligned}$$

I'll assume Beauty is certain that if she wakes twice, she will make the same bets both times. This assumption is highly plausible. Beauty's epistemic situation on when she is offered a bet Monday is exactly like her situation when she is offered a bet on Tuesday. Presumably the same dispositions that influence her Monday decision also influence her Tuesday decision. Using that assumption, we can simplify the above expected value calculations:

$$\begin{aligned} V_E(\textit{Accept}) \quad &= Cr(\textit{Heads}|\textit{Accept})V(\textit{Heads} \,\&\, \neg D \,\&\, \textit{Accept}) \\ &\quad + Cr(\textit{Tails}|\textit{Accept})V(\textit{Tails} \,\&\, D \,\&\, \textit{Accept}) \\ &= Cr(\textit{Heads}|\textit{Accept})\$(10+2\epsilon) \\ &\quad + Cr(\textit{Tails}|\textit{Accept})\$(-20+2\epsilon) \end{aligned}$$

$$\begin{aligned} V_E(\neg \textit{Accept}) \quad &= Cr(\textit{Heads}|\neg \textit{Accept})V(\textit{Heads} \,\&\, \neg D \,\&\, \neg \textit{Accept}) \\ &\quad + Cr(\textit{Tails}|\neg \textit{Accept})V(\textit{Tails} \,\&\, \neg D \,\&\, \neg \textit{Accept}) \\ &= \$0 \end{aligned}$$

Furthermore, it seems reasonable to assume that Beauty's acceptance or rejection of a bet gives her no evidence as to the outcome of the coin toss—in other words, that $Cr(Heads|Accept) = Cr(Heads)$. Thus,

$$V_E(Accept) = Cr(Heads)\$(10+\epsilon) + Cr(Tails)\$(-20+2\epsilon)$$

If Beauty is a halfer (that is, if $Cr(Heads) = 1/2$), it turns out that $V_E(Accept) = \$(-5 + 3\epsilon/2)$. Since ϵ is minute, $\$(-5 + 3\epsilon/2) < \$0$, and so $V_E\ (Accept) < V_E\ (\neg Accept)$. So if Beauty is both a halfer and an evidential decision theorist, she should reject bet 10, and Hitchcock's Dutch book argument fails.

On the other hand, if Beauty is a halfer and a causal decision theorist, she should accept bet 10. *Heads*, *Tails* & *D* and *Tails* & $\neg D$ are paradigmatic dependency hypotheses. Each of these propositions determines which outcome Beauty's choice will cause, but Beauty's choice has no causal impact on which of them is true. So the causal expected values of *Accept* and $\neg$*Accept* are

$$\begin{aligned} V_C(Accept) &= Cr(Heads)V(Heads\,\&\,Accept) \\ &\quad + Cr(Tails\,\&\,D)V(Tails\,\&\,D\,\&\,Accept) \\ &\quad + Cr(Tails\,\&\,\neg D)V(Tails\,\&\,\neg D\,\&\,Accept) \\ &= (1/2)\$(10+\epsilon) \\ &\quad + Cr(Tails\,\&\,D)\$(-20+\epsilon) \\ &\quad + Cr(Tails\,\&\,\neg D)\$(-10+\epsilon) \end{aligned}$$

$$\begin{aligned} V_C(\neg Accept) &= Cr(Heads)V(Heads\,\&\,\neg Accept) \\ &\quad + Cr(Tails\,\&\,D)V(Tails\,\&\,D\,\&\,\neg Accept) \\ &\quad + Cr(Tails\,\&\,\neg D)V(Tails\,\&\,\neg D\,\&\,\neg Accept) \\ &= Cr(Tails\,\&\,D)\$(-10+\epsilon) \end{aligned}$$

No matter what the value of *Cr*(*Tails* & *D*) is,

$$\begin{aligned} V_C(Accept) - V_C(\neg Accept) &= 1/2\$(10+\epsilon) \\ &\quad + Cr(Tails\,\&\,D)\$(-10+\epsilon) \\ &\quad + Cr(Tails\,\&\,\neg D)\$(-10+\epsilon) \\ &= \$\epsilon \end{aligned}$$

In other words, accepting the bets has a slightly higher causal expected value than rejecting them. If Beauty is a causal decision theorist as well as a halfer, she should accept bet 10, and she is vulnerable to Hitchcock's Dutch book.

The moral of this section is that in cases of irreducibly *de se* ignorance, an an agent's credences do not always match her betting odds—at least if she is an evidential decision theorist. In examples like *Sleeping Beauty*, accepting a bet at one center is *correlated* with gains or losses at other centers, although it does not *cause* gains or losses at other centers. (When an agent does not suffer from irreducibly *de se* ignorance, we can expect her betting odds to match her credences whether or not she is an evidential decision theorist, since she bets at only one center per world.)

The divergence between credences and betting odds answers an outstanding question from section 2.1. Why is Beauty vulnerable to a Dutch book if she satisfies equation 5, but not necessarily if she satifies equation 4? In formulating their Dutch book, Draper and Pust (2008) assume that Beauty's betting odds match her credences. This assumption is reasonable after Beauty learns that it's Monday, since she has only one center per doxastically possible uncentered world. It is not reasonable when she wakes up, since she has more than one center per doxastically possible uncentered world.

3.3. *A New Dutch Book*

In the previous section, I showed that that causal decision theorists who obey the Halfer Rule fall victim to Hitchcock's Dutch book. Evidential decision theorists who obey the Thirder Rule are vulnerable to a Dutch book of their own. Consider the following set of bets, the first of which takes place on Sunday:

$$\begin{array}{ll} Heads & \$(15+2\epsilon) \\ Tails & \$(-15+\epsilon) \end{array} \tag{11}$$

and the second of which takes place on Monday and (and on Tuesday, if Beauty is awake):

$$\begin{array}{ll} Heads & \$(-20+\epsilon) \\ Tails & \$(5+\epsilon) \end{array} \tag{12}$$

If Beauty accepts bets 11 and 12, she is bound to lose $\$(5-3\epsilon)$. If the coin lands heads, she wins $\$(15+2\epsilon)$ on bet 11, but loses $\$(20-\epsilon)$ on bet 12. If it lands tails she loses $\$(15-\epsilon)$ on bet 11, and wins back only $\$(10+2\epsilon)$ on bet 12.

If Beauty is a thirder and an evidential decision theorist, then she should accept bets 11 and 12. It is clear that on Sunday, the expected value of accepting bet 11 is greater than the expected value of refusing it by $\$3\epsilon/2$. When Beauty wakes and considers bet 12, where *Accept* is the centered proposition that she accepts bet 12 right now, and *D* is the proposition that she accepts bet 12 on a different day, the evidential expected values of *Accept* and $\neg$*Accept* are as follows:

$$\begin{aligned} V_E(Accept) &= Cr(Heads|Accept)V(Heads\,\&\,\neg D\,\&\,Accept) \\ &\quad +Cr(Tails|Accept)V(Tails\,\&\,D\,\&\,Accept) \\ &= (1/3)(-20+\epsilon) + (2/3)(10+2\epsilon) \\ &= 3\epsilon \end{aligned}$$

$$\begin{aligned} V_E(\neg Accept) &= Cr(Heads|\neg Accept)V(Heads\,\&\,\neg D\,\&\,\neg Accept) \\ &\quad +Cr(Tails|\neg Accept)V(Tails\,\&\,\neg D\,\&\,\neg Accept) \\ &= 0 \end{aligned}$$

In other words, for evidential decision theorists who obey the Thirder Rule, bets 11 and 12 constitute a Dutch book.

So far, I've proved two Dutch book results: causal decision theorists who obey the Halfer Rule are vulnerable to Hitchcock's Dutch book, and evidential decision theorists who obey the Thirder Rule are vulnerable my Dutch book. In the next two sections, I will prove two converse Dutch book results: causal decision theorists who obey the Thirder Rule and evidential decision theorists who obey the Halfer Rule are immune to Dutch books. Section 3.4 proves a lemma which is crucial to both results: agents who bet at what I call *thirder odds* are immune to Dutch books. Section 3.5 proves that (given some reasonable background assumptions) causal decision theorists who obey the Thirder Rule and evidential decision theorists who obey the Halfer Rule bet at thirder odds.

3.4. *Why Everyone Should Bet at Thirder Odds*

In order to define thirder odds, I must first stipulate what counts as a betting arrangement. Let a betting arrangement be a partition A that divides uncentered worlds into equivalence classes such that the agent receives the same payoff at any two centers belonging to uncentered worlds in the same equivalence class. I'll switch back and forth between writing agent's payoff at world W as X_W and writing the agent's payoff in an equivalence class A as X_A.

An agent bets at *thirder odds* for an uncentered credence function Cr_u just in case she is willing to accept all and only betting arrangements such that, where W is a variable ranging over uncentered worlds,

$$\sum_W N_W Cr_u(W) X_W > 0$$

Halfer odds will play an important role in my argument. I'll argue that an agent bets at halfer odds for Cr_u just in case for every uncentered world W, she is willing to accept all and only betting arrangements such that

$$\sum_W Cr_u(W) X_W > 0$$

(Intuitively, thirder odds are odds that 'correspond' to thirder credences, and halfer odds are odds that 'correspond' to halfer credences.) I will argue that agents who bet at thirder odds are immune to Dutch books (provided Cr_u is coherent and updated by conditionalization).

The bare bones of my argument are as follows. Suppose an agent (called 'the original agent' for reasons that will soon become apparent) has a coherent uncentered credence function Cr_u, and suppose she bets at thirder odds for Cr_u at all of her doxastically possible centers. We can describe a second, imaginary agent who bets at halfer odds for Cr_u, and who bets only once at each uncentered world. Teller (1973) has proved that (as long as Cr_u is updated by conditionalizing) such an imaginary agent is immune to Dutch books. But for any betting arrangement that the original agent is willing to accept, we can generate a corresponding bet β' such that

(a) the imaginary agent is willing to accept β' and
(b) for every $A \in A$ such that $X_A \neq 0$, there is some $W \in A$ such that at W, the original agent wins at least as much as the imaginary agent.

If we could subject the original agent to a Dutch book, then we could subject the imaginary agent to a Dutch book by selling her the corresponding bets before and after she updated. But by Teller's result, the imaginary agent is immune to Dutch books. Therefore, the original agent must be immune to Dutch books too.

The bare bones need fleshing out in a few places. First, I must specify a procedure for generating the imaginary agent's betting arrangement β' from the original agent's betting arrangement β. Second, I must show that (a)–(c) hold true for β'. I'll take these tasks in turn.

Let the expected number of centers in $A \in A$ be defined as

$$C_A = \frac{\sum_{W \in A} Cr_u(W) N_W}{Cr_u(A)}$$

Let β' be a betting arrangement that pays, for any $A \in \mathcal{A}$,

$$\$X'_A =_{df} \$C_A X_A$$

Now that I've defined β', it only remains to show (a) and (b).

I'll begin with (a). Suppose the original agent is willing to accept β. Since she bets at thirder odds,

$$\sum_W N_W Cr_u(W) X_W > 0$$

$$\sum_{A \in \mathcal{A}} \sum_{W \in A} N_W Cr_u(W) X_A > 0$$

$$\sum_{A \in \mathcal{A}} \frac{\sum_{W \in A} Cr_u(W) N_W}{Cr_u(A)} X_A Cr_u(A) > 0$$

$$\sum_{A \in \mathcal{A}} C_A X_A Cr_u(A) > 0$$

$$\sum_W X'_W Cr_u(W) > 0$$

Since the imaginary agent bets at halfer odds, it follows that she'll be willing to accept β. β' satisfies (a), and it only remains to show that β satisfies (b).

For $X_A > 0 (X_A < 0)$, and let α_A be a world in A with at least as many (few) centers as any other world in A. The maximum number of centers in A is at least as large as the expected number of centers in A, so $X_{\alpha A} \geq X'_{\alpha A}$. β' satisfies (b).

3.5. *Who Bets at Thirder Odds?*

I've shown that agents who bet at thirder odds are immune to Dutch books. In this section, I'll show that causal decision theorists who satisfy the Thirder Rule and evidential decision theorists who satisfy the Halfer Rule are immune to Dutch books. My argument will require a few assumptions.

First, I'll assume that any agent who suffers from irreducibly *de se* ignorance is certain that within each uncentered world, she will bet the same way at every subjectively indistinguishable center. I gave a reason for accepting this assumption in section 3.2: at any two subjectively indistinguishable centers in an uncentered world, an agent's beliefs and desires are the same (else she would be able to distinguish between the two centers by taking note of her beliefs and desires). The agents I'll be discussing are either consistent causal decision theorists or consistent evidential decision theorists, so at any two subjectively indistinguishable centers in an uncentered world, they use the same decision procedure. If an agent's choices are determined by her beliefs, desires, and decision procedure, then she will have to bet the same way at any subjectively indistinguishable centers in an uncentered world.

Second, I'll assume that an agent's accepting or refusing bets gives her no evidence about the propositions she is betting on. Cases that violate this assumption are problematic for any proponent of Dutch books, not just for me, and a full discussion of such cases would be a distraction from the topic at hand. So, I will set them aside.

Third, I'll assume that accepting bets does not cause the agent to gain or lose anything, aside from what she wins or loses as the bet's explicit payoff. I take this to be a stipulation about what counts as a bet.

When the above three assumptions are satisfied, causal decision theorists who obey the Thirder Rule are invulnerable to Dutch books. For suppose there is a causal decision theorist who obeys the Thirder Rule, and suppose she is considering a betting arrangement β.

We know that for each $A \in \mathcal{A}$ if A is the case, then no matter what else is true, the agent will be $\$X_A$ richer if she accepts β than if she rejects it. Therefore, the causal expected value of accepting β is

$$V_C(Accept) = \$\sum_{A \in \mathcal{A}} X_A Cr_@(A)$$

By the Thirder Rule,

$$V_C(Accept) = \$\sum_{A \in \mathcal{A}} X_A \frac{\sum_{W \in A} N_W Cr_u(W)}{\sum_{W*} N_W Cr_u(W*)}$$

The agent, being a causal decision theorist, will accept the bet only when $V_C(Accept) > 0$. This will happen just in case

$$\sum_{A \in \mathcal{A}} X_A \frac{\sum_{W \in A} N_W Cr_u(W)}{\sum_{W_*} N_W Cr_u(W_*)} > 0$$

Or, since $\sum_{W*} N_W Cr_u(W^*) > 0$, just in case

$$\sum_{A \in \mathcal{A}} X_A \sum_{W \in A} N_W Cr_u(W) > 0$$

$$\sum_{W \in \mathcal{A}} N_W Cr_u(W) X_W > 0$$

I've proved that the agent bets at thirder odds.

Likewise, an evidential decision theorist who obeys the Halfer Rule will bet at thirder odds. For suppose there is an evidential decision theorist who obeys the Halfer Rule, and suppose she is considering a betting arrangement β. If the agent occupies an uncentered world with N_W centers, then a centered world where she accepts β will have value $\$N_W X_W$ (since if she accepts the bet on this occasion, she is guaranteed to accept it again on $N_W - 1$ other occasions). The expected value of accepting β is:

$$V_E(Accept) = \$\sum_W Cr_@(W) N_W X_W$$

By the Halfer Rule,

$$V_E(Accept) = \$\sum_W Cr_u(W) N_W X_W$$

Since the agent is an evidential decision theorist, she will accept β if and only if $V_E(Accept) > 0$. But to accept a betting setup if and only if $\sum_W Cr_u(W) N_W X_W > 0$ is just to bet at thirder odds.

4. SCORING RULES

Kierland and Monton (2005) suggest that we use scoring rules to adjudicate the debate between halfers and thirders. I will follow Kierland and Monton in focusing on the Brier score. It's not obvious that the Brier score is the only acceptable method of measuring inaccuracy, but there are reasons to expect that what holds for the Brier score holds for scoring rules generally.

In section 4.1, I'll explain how the Brier score works as a measure of inaccuracy, discuss Kierland and Monton's suggestion that agents ought to minimize expected inaccuracy, and criticize Kierland and Monton's definition of expected inaccuracy. In section 4.2, I'll suggest two ways of revising the definition of expected inaccuracy. In section 4.3, I'll argue causal decision theorists should favor one definition of expected inaccuracy, while evidential decision theorists should favor the other. According to the definition that's best for causal decision theorists, the Thirder Rule minimizes expected inaccuracy. According to the definition that's best for evidential decision theorists, the Halfer Rule minimizes expected inaccuracy. Once again, it turns out that causal decision theorists should obey the Thirder Rule, while evidential decision theorists should obey the Halfer Rule.

4.1. Measuring Inaccuracy

The idea behind scoring rules is something like this: just as full belief aims at truth, partial belief aims at accuracy. If A is true, it's good to fully believe A and bad to fully believe $\neg A$, while if A is false, it's good to fully believe A and bad to fully believe A. Likewise, if A is true, it's is good to believe A to a high degree (the higher the better) and bad to believe A to a low degree (the lower the worse), while if A is false, then it's good to believe A to a low degree (the lower the better) and bad to believe A to a high degree (the higher the worse). If it turns out that halfing (or thirding) conduces to inaccuracy, this will give us a reason to reject halfing (or thirding).

In addition to *ranking* partial beliefs in terms of inaccuracy, we can *measure* their inaccuracy. Among the many possible measures, the *Brier score* (developed by Brier (1950)) is the most popular. Where X is a *de dicto* proposition, $Cr_@(X)$ is an agent's credence in X, and $W(X)$ is X's truth value, the agent's Brier score for X is

$$S_W(X) = (W(X) - Cr_@(X))^2 \qquad (13)$$

The lower the Brier score—that is, the closer the agent's degree of belief in A to A's truth value—the better.

In addition to measuring inaccuracy for individual partial beliefs, we can use the Brier score to measure inaccuracy for sets of partial beliefs. Where $\{X_1, X_2, \ldots X_n\}$ are (*de dicto*) propositions in the domain of the agent's credence function, $W(X_i)$ is X_i's truth value, and $Cr_@(X_i)$ is the agent's credence in X_i, the agent's expected inaccuracy for $\{X_1, X_2, \ldots X_n\}$, as calculated using the Brier score, is:

$$S_W(X_1, X_2, \ldots X_n) = \sum_i S_W(X_i) \qquad (14)$$

An important advantage of the Brier score is that it's a *proper* scoring rule, meaning that an agent whose goal is to minimize her Brier score will have no incentive to 'cheat' by adjusting her credences in ways that are epistemically unwarranted. Savage (1971) shows that if value is defined in terms of accuracy, so that the value of adopting a set of partial beliefs $\{X_1, X_2, \ldots X_n\}$ is $-S_W$ $(X_1, X_2, \ldots X_n)$, then an agent will always assign higher expected value to keeping her current beliefs about $\{X_1, X_2, \ldots X_n\}$ than to adopting any other set of beliefs about $\{X_1, X_2, \ldots X_n\}$. Savage uses the evidential definition rather than the causal definition of expected value, but in ordinary cases, we shouldn't expect this to make much difference. As long as the agent's adopting a set of beliefs is uncorrelated with her accumulating inaccuracy by some other means, the two definitions will coincide. The Brier score is not the only proper scoring rule, but what I have to say about the Brier score should apply to any other proper scoring rule.

Advising a partial believer to minimize her inaccuracy isn't very helpful. Someone who is unsure what the world is like won't know how inaccurate her partial beliefs are. But if she knows the objective chances of the world's turning out various ways, she can calculate her expected inaccuracy. Let $\{W_1, W_2, \ldots W_m\}$ be a set of 'possible worlds', or logically consistent valuation functions assigning truth values to each of $\{X_1, X_2, \ldots X_n\}$. Then where $\{X_1, X_2, \ldots X_n\}$ are uncentered propositions and $P(X_i)$ is the objective chance of X_i, we can define the agent's expected inaccuracy for $\{X_1, X_2, \ldots X_n\}$ as follows:

$$S_E(X_1, X_2, \ldots X_n) = \sum_j P(W_j) \sum_i S_{W_j}(X_i)$$

Kierland and Monton 2005 suggest that when an agent knows the relevant objective chances, she should minimize her expected inaccuracy.

This suggestion is on the right track, but there are three potential problems with it. First an agent may have what Lewis (1986) calls 'inadmissible information'—information that tells her more about whether an event will occur than the event's objective chance alone. To take a somewhat farfetched example, suppose an agent has a single atom of carbon 14, which a trustworthy oracle tells her will decay within three months. On the basis of the oracle's testimony, it may be reasonable for her to place a high credence in the proposition that the atom decays within three months, even thought the objective chance of its decaying so quickly is low. When an agent has inadmissible information, it looks like minimizing expected inaccuracy is the wrong thing to do.

Second, the advice to minimize expected inaccuracy is not invariably useful (as Kierland and Monton point out). Agents don't always know the objective chances. If an agent doesn't know the value of $P(W_j)$ for some W_j generated by $\{X_1, X_2, \ldots X_n\}$, then she won't know how to minimize $S_E(X_1, X_2, \ldots X_n)$.

The third problem is a bit more subtle than the first two. So far, my talk of 'propositions' and 'worlds' has been ambiguous. Objective chances attach to uncentered worlds, but the agent's credences attach to centered worlds. So the W_j in $P(W_j)$ can't be the same sort of thing as the W_j in $S(W_j)$. The ambiguity is harmless in cases of purely *de dicto* ignorance, where the agent's epistemic state can be represented in terms of uncentered worlds. But in cases of irreducibly *de se* ignorance, there is no clear way of paraphrasing the definition of expected inaccuracy to eliminate the ambiguity.

On Kierland and Monton's definition of expected inaccuracy, the advice to minimize expected inaccuracy is sometimes wrong and sometimes useless. Furthermore, in cases of irreducibly *de se* ignorance, there's no univocal way of interpreting the definition. These are the problems—how best to remedy them?

4.2. *Revising the Concept of Expected Inaccuracy*

For my current purposes (adjudicating between the Halfer Rule and the Thirder Rule), there's a convenient fix for the first two problems. We can define expected inaccuracy not in terms of the objective chance function, but in terms of the agent's uncentered credence function Cr_u. Just as Cr_u ought to satisfy conditionalization, it ought to satisfy the Principal Principle—$Cr_u(A|P(A) = x)$ should equal x, provided the agent has no admissible information. And if an agent satisfies the Principal Principle, then if she knows A's objective chance and has no inadmissible information, $Cr_u(A)$ will equal $P(A)$.

So when a definition of expected inaccuracy in terms of the objective chance function gets the right answers, a definition in terms of Cr_u gets the right answers too. When a definition of expected inaccuracy in terms of the objective chance function gets the wrong answers (because the agent has inadmissible information) a definition in terms of Cr_u can do better, because Cr_u takes the inadmissible information into account. And when definition of expected inaccuracy in terms of the objective chance function yields no useful answers (because the agent has too little information about the objective chances), a definition in terms of Cr_u can do better, because the agent presumably has better epistemic access to Cr_u than to the objective chance function.

The third problem is not yet solved—like P, Cr_u assigns probabilities only to uncentered propositions. We can solve the third problem by letting the W_js be centered worlds, and letting $u(W_j)$ designate the uncentered world associated with W_j. Instead of calculating expected inaccuracy using $Cr_u(W_j)$, which is undefined, we can calculate it using $Cr_u(u(W_j))$.

We're almost ready to introduce a new, improved definition of expected inaccuracy. There's just one problem: the correspondence between centered worlds and uncentered worlds is not one–one, but many–one. There are two natural ways extending the definition of expected inaccuracy. (These ideas are faintly suggested by the calculations in Kierland and Monton (2005), though the authors don't comment very extensively on the general principles behind the calculations. Nonetheless, I will follow Kierland and Monton's terminology.)

We might measure expected inaccuracy as expected total inaccuracy:

$$S_{ET}(X_1, X_2, \ldots X_n) = \sum_j Cr_u(u(W_j)) \sum_i S_{W_j}(X_i)$$

On the other hand, we might measure expected inaccuracy as expected average inaccuracy:

$$S_{EA}(X_1, X_2, \ldots X_n) = \sum_j Cr_u(u(W_j)) \sum_i \frac{\sum_i S_{W_j}(X_i)}{N_{W_j}}$$

Both measures look plausible. Expected total inaccuracy is uniform across centered worlds—each centered world makes the same contribution to S_{EA}, proportional to the probability assigned to the corresponding centered world by Cr_u. Expected average inaccuracy is uniform across uncentered worlds—each uncentered world makes the same contribution to S_{EAT}, proportional to the probability assigned to it by Cr_u. Which measure is best? And what does the best measure have to tell us about irreducibly *de se* ignorance?

4.3. *Average or Total? Halfer or Thirder?*

I pointed out in section 4.1 that the Brier score had the advantage of being a proper scoring rule. A good and important thing about proper scoring rules is that they don't generate pragmatic rewards for epistemically unwarranted credences.

One might worry that choosing the wrong measure of expected inaccuracy will generate pragmatic rewards for epistemically unwarranted credences, in something like the way an improper scoring rule does. In particular, it would be problematic if a measure of expected inaccuracy weighted an agent's inaccuracy more heavily at some centered worlds than at others.

Does either S_{EA} or S_{ET} have a problem with weighting? This depends on what the 'penalty' for inaccuracy is. In one sense—a causal sense—the agent's inaccuracy at a center is independent of the number of subjectively indistinguishable centers in the same uncentered world. Her having accurate or inaccurate beliefs at one center does not cause her to have accurate or inaccurate beliefs at subjectively indistinguishable centers in the same world. If we understand 'reward' causally, the agent's inaccuracy at each center should count for the same amount. Otherwise, she will be rewarded for reporting credences that are too low for worlds with multiple centers (since her accuracy at each of the centers will count for less). So on the causal understanding of 'reward', S_{ET} seems to be the better scoring rule.

In an evidential sense, on the other hand, the agent's inaccuracy at a center depends not only on what she believes at that center, but also on the number of subjectively indistinguishable centers in the same world. Her having accurate or inaccurate beliefs at one center is decisive evidence that she has equally accurate or inaccurate beliefs at other, subjectively indistinguishable centers.

If we understand 'reward' in this evidential sense, the agent's inaccuracy at each uncentered world should count for the same amount, no matter how many centers the uncentered world contains. Otherwise, she will be rewarded for reporting credences that are too high for worlds with multiple centers (since her accuracy at each of the centers will count for more). On the evidential understanding of 'reward', S_{EA} seems to be the better scoring rule.

So causal decision theorists should minimize expected total inaccuracy, while evidential decision theorists should minimize expected average inaccuracy. It will turn out that agents who aim to minimize expected total inaccuracy should adhere to the Thirder Rule, while agents who aim to minimize expected average inaccuracy should adhere to the Halfer Rule.

Suppose an agent wants to now how much credence to place in an individual world *W*, in order to minimize the expected total inaccuracy of her beliefs regarding *W*. Then she'll want to minimize

$$\begin{aligned} S_{ET}(W) &= Cr_u(W)N_W(1 - Cr_@(W))^2 + \sum\nolimits_{W^* \neq W} Cr_u(W_*)N_{W^*}Cr_@(W)^2 \\ &= Cr_u(W)N_W(1 - 2Cr_@(W) + (Cr_@(W)^2) \\ &+Cr_u(W^*)N_{W_*}Cr_@(W)^2 \end{aligned}$$

We can find the minimum of this equation by setting its derivative with respect to $Cr_@(W)$ equal to zero.

$$\begin{aligned} &Cr_u(W)N_W(-2 + 2Cr_@(W)) + 2(\sum\nolimits_{W^*} Cr_u(W^*)N_{W^*})Cr_@(W) = 0 \\ &-2Cr_u(W)N_W + 2Cr_@(W)(Cr_u(W)N_W + (\sum\nolimits_{W^*} Cr_u(W^*)N_{W^*}) = 0 \\ &-2Cr_u(W)N_W + 2Cr_@(W)(\sum\nolimits_{W^*} Cr_u(W^*)N_{W*}) = 0 \\ &Cr_u(W)N_W = Cr_@(W)(\sum\nolimits_{W^*} Cr_u(W^*) \\ &Cr_@(W)(\sum W^* Cr_u(W^*) = Cr_u(W)N_W \\ &Cr_@(W) = \frac{Cr_u(W)N_W}{\sum W^* Cr_u(W^*} \end{aligned}$$

The upshot: if an agent is to minimize the expected total inaccuracy of her credence in a proposition consisting of a single uncentered world *W*, her credence in that proposition must accord with the thirder rule.

Suppose an agent wants to be now how much credence to place in an individual world *W*, in order to minimize, not the expected total inaccuracy, but the expected average inaccuracy of her beliefs regarding *W*.

$$\begin{aligned} S_{EA}(W) &= Cr_u(W)N_W\frac{(1 - Cr_@(W))^2}{N_W} + \sum\nolimits_{W_* \neq W} Cr_u(W^*)N_{W^*}\frac{Cr_@(W)^2}{N^*_W} \\ &= Cr_u(W)(1 - Cr_@(W))^2 + \sum\nolimits_{W_* \neq W} Cr_u(W^*)Cr_@(W)^2 \\ &= Cr_u(W)(1 - 2Cr_@(W) + (Cr_@(W))^2) + \sum\nolimits_{W_* \neq W} Cr_u(W^*)Cr_@(W)^2 \end{aligned}$$

We can find the minimum of this equation by setting its derivative with respect to $Cr_@(W)$ equal to zero:

$$\begin{aligned} &Cr_u(W)(-2 + 2Cr_@(W)) + \sum\nolimits_{W^* \neq W} Cr_u(W^*)2Cr_@(W) = 0 \\ &-2Cr_u(W) + 2Cr_@(W)(Cr_u(W) + \sum\nolimits_{W^* \neq W} Cr_u(W^*)) = 0 \\ &-2Cr_u(W) + 2Cr_@(W) = 0 \\ &Cr_u(W) = Cr_@(W) \end{aligned}$$

So an agent's expected average inaccuracy for an uncentered world has its minimum when $Cr_@ = Cr_u$—in other words, when she satisfies the Halfer Rule.

The Thirder Rule furthers the goal of minimizing expected total inaccuracy (the criterion that causal decision theorists should favor), while the Halfer Rule furthers the goal of minimizing expected average inaccuracy (the criterion that evidential decision theorists should favor). Again, causal decision theory goes with the Thirder Rule, and evidential decision theory goes with the Halfer Rule.

5. STABILITY

There is one criterion that favors the Thirder Rule over the Halfer Rule independently of any decision-theoretic considerations. An adequate *de se* confirmation theory should be insensitive to relatively trivial changes in the way we represent an agent's credence function. In particular, an adequate *de se* confirmation theory should be stable in the following sense: if we care about changes in an agent's credence regarding *A*, then once the agent's doxastic worlds are represented richly enough to include all information she considers relevant to *A*, it should be possible to enrich them without changing the theory's advice.

We can illustrate the concept of stability using conditionalization. Suppose that we are interested in the effects of conditionalizing on an agent's beliefs regarding some proposition *A*. And suppose that, according to our way of representing the agent's credence function, *E* is the strongest proposition of which the agent becomes certain between between t_0 and t_1. There might be some irrelevant proposition *B* such that the agent becomes certain of *B* between t_0 and t_1, but the belief worlds in our representation need not settle whether *B* is true or false. We can enrich our representation, dividing each of the agent's belief worlds into one world where *B* is true, and another where *B* is false. Since *B* is irrelevant to *A*, we should ensure that at t_0 $Cr_0(A|E \,\&\, \mathrm{B}) = Cr_0(A|E)$. Enriching our representation this way won't affect conditionalization's advice regarding the proposition *A*.

The following simple example shows how conditionalizing is a stable rule: Linda is about to witness a fair coin toss. Before the toss, she grants credence 1/2 to the proposition that the coin is fair and credence 1/2 to the proposition that it is biased with *P(Heads)* = ¼. She then learns that the coin is fair, and updates accordingly. Of course, while learning that the coin is fair, she gains plenty of information which is irrelevant to the outcome of the coin toss—that the coin's owner told her the coin was fair, that he told her in a pleasant baritone voice, and so forth. We might represent Linda's initial credence function Cr_0 as follows, where *Fair* is the proposition that the coin is fair, and *Biased* is the proposition that it is biased:

	Heads	Tails
Fair	1/4	1/4
Biased	1/8	3/8

(15)

On this way of representing things, Linda conditionalizes on *Fair*, so that $Cr_1(\mathit{Heads}) = 1/2$.

On the other hand, we might represent Cr_0 in a somewhat richer way, where B is the proposition that Linda learns about the coin's bias from someone with a nice baritone voice. (Since B is irrelevant information, it must be the case that $P(\mathit{Heads}|\, B \,\&\, \mathit{Fair}) = P(\mathit{Heads}|\mathit{Fair})$).

	Heads	Tails
Fair & B	1/16	1/16
Fair & ¬B	3/16	3/16
Biased & B	1/32	3/32
Biased & ¬B	3/32	9/32

(16)

On this richer way of representing things, Linda conditionalizes on *Fair & B*, and once again, $Cr_1(\mathit{Heads}) = 1/2$.

There are numerous other ways of representing Linda's credence function, corresponding to different ways of representing the irrelevant information that Linda might learn. But as long as we make sure that E screens A off from any irrelevant information, then any way of representing Linda's credence function, when taken together with conditionalization, yields the result that $Cr_1(\mathit{Heads}) = 1/2$.

We might enrich our representation of SB's belief worlds, just as we enriched our representation of Linda's belief worlds. Titelbaum (forthcoming) illustrates this idea using the following variant on the Sleeping Beauty story, which he calls *Technicolor Beauty*:

> Everything is exactly as in the original Sleeping Beauty Problem, with one addition: Beauty has a friend on the experimental team, and before she falls asleep Sunday night he agrees to do her a favor. While the other experimenters flip their fateful coin, Beauty's friend will go into another room and roll a fair die. (The outcome of the die roll is independent of the outcome of the coin flip.) If the die roll comes out odd, Beauty's friend will place a piece of red paper where Beauty is sure to see it when she awakens Monday morning, then replace it Tuesday morning with a blue paper she is sure to see if she awakens on Tuesday. If the die roll comes out even, the process will be the same, but Beauty will see the blue paper on Monday and the red paper if she awakens on Tuesday.
>
> Certain that her friend will carry out these instructions, Beauty falls asleep Sunday night. Some time later she finds herself awake, uncertain whether it is Monday or Tuesday, but staring at a colored piece of paper. What does ideal rationality require at that moment of Beauty's degree of belief that the coin came up heads?

Technicolor Beauty is just an enriched version of *Sleeping Beauty*. Seeing the piece of paper tells Beauty nothing about the outcome of the coin toss—she is equally likely to see a red piece of paper whether the coin lands heads or tails. Likewise, seeing the piece of paper tells Beauty nothing about what day it is—she is equally likely to see a red piece of paper whether it is Monday or Tuesday. Whatever Beauty ought to do in *Sleeping Beauty*, she ought to do in *Technicolor Beauty*. The only difference between the two scenarios is that in *Technicolor Beauty*, Beauty gains irrelevant *de dicto* information upon waking up.

Although the Halfer Rule and the Thirder Rule disagree in *Sleeping Beauty*, they agree in *Technicolor Beauty*. I'll assume that on Sunday night in the *Technicolor Beauty* example, Beauty assigns credence 1/4 to each of the *de dicto* propositions *Heads & Even*, *Heads & Odd*, *Tails & Even*, and *Tails & Odd*, where *Even* and *Odd* designate the relevant outcomes of the die roll. Upon waking up and seeing the paper, Beauty eliminates either *Heads & Even* (if she sees red paper), or *Heads & Odd* (if she sees blue paper). In either case, if Beauty updates Cr_u by conditionalizing on the *de dicto* portion of her total evidence, then

$$Cr_u(\mathit{Heads}) = 1/3$$

But once Beauty has seen the paper, there is only one center in each of her doxastically possible uncentered worlds. Knowing the outcome of the coin toss and the die toss would be enough to tell her what day it is. Therefore, the Halfer Rule and the Thirder Rule agree that $Cr_{@}(\mathit{Heads}) = Cr_u(\mathit{Heads}) = 1/3$.

Enriching Beauty's doxastically possible centered worlds with irrelevant information has changed the Halfer Rule's advice. In *Sleeping Beauty*, the Halfer Rule recommended that $Cr^{up}(\mathit{Heads}) = 1/2$, but in *Technicolor Beauty*, it recommends that $Cr^{up}(\mathit{Heads}) = 1/3$. The Halfer Rule is unstable. On the other hand, enriching Beauty's doxastically possible worlds with irrelevant information has not effected the Thirder Rule's advice. In both *Sleeping Beauty* and *Technicolor Beauty*, it recommends that that $Cr^{up}(\mathit{Heads}) = 1/3$. So the Thirder Rule is stable—at least in this particular case.

Notice that Beauty's friend is not essential to the point Titelbaum is making in his example. What's crucial is that Beauty be convinced that she will have a different experience every time she wakes up. Beauty can convince herself of this point by putting a coin in her pocket on Sunday night, flipping it ten times whenever she wakes up, and observing the string of outcomes.[4] If she follows

[4] I thank Robert Stalnaker for pointing this out to me.

this coin-flipping procedure, her chance of getting the same string of outcomes on both days is tiny—less than .0005. The Halfer Rule will instruct her to set her credence in *Heads* close to 1/3 upon waking up and tossing the coin.

The essential problem is that the distinction between purely *de dicto* ignorance and irreducibly *de se* ignorance is sensitive to the way an agent's situation is represented. Enrich the agent's doxastically possible uncentered worlds, and you can change a purely *de se* matter to a *de dicto* matter. The Halfer Rule, which instructs agents to ignore the *de se* portion of their total evidence, gives different advice depending on how richly we represent each doxastically possible uncentered world.

The Thirder Rule, on the other hand, is insensitive to such changes. It is stable in a more general sense, which can be cashed out formally. Suppose that (relative to some representation) an agent suffers from irreducibly *de se* ignorance at time t_1. Let B be an irrelevant centered proposition which the agent learns at t_1 by direct observation. (I will establish a precise sense of "irrelevant" in the next paragraph. But from an intuitive standpoint, I have in mind the sort of centered proposition expressed by sentences like "I am looking at a red piece of paper now," "The birds are singing loudly now," or "I am now having sui generis subjective experience S that cannot be duplicated.") Suppose that on our original representation of the agent's credence function, none of her doxastically possible centered worlds settles whether B. I claim that if we enrich the agent's doxastically possible centered worlds so that each settles whether B, and we correspondingly enrich her doxastically possible uncentered worlds so each settles the question of which centers B is true at, this will not affect the Thirder Rule's prescriptions.

What does it mean, in more precise terms, to say that B is irrelevant? I will define a formal property that I think captures the idea. Assume that there is a time t_0, shortly before t_1, such that according to the original representation, the agent's total uncentered evidence is the same at t_0 and t_1, but according to the enriched representation, the agent does not know at t_0 whether B will ever obtain. We can then say that B is irrelevant in the formal sense if at t_0 there is some real number δ such that for any center which will be doxastically possible at t_1 in any uncentered world W of the original representation, the agent at t_0 places credence δ in B's obtaining conditional on W; and if in addition, one of the following conditions holds:

(a) B is unrepeatable: where W_1 and W_2 are centers in the same uncentered world, the agent's t_0 credence in B's obtaining at W_1 conditional on B's obtaining at W_2 is 0, or

(b) occurrences of B are independent: according to the agent's t_0 credences, B is binomially distributed among the centers in each original uncentered world that will be doxastically possible at t_1.

We can model the formal property of irrelevance by supposing that each world corresponds to an urn filled with some large finite number of balls, one of which is labeled B. (All the urns, we assume, contain the same number of balls.) We can imagine that at a doxastically possible uncentered world in the original representation, each center corresponds to a trial in which one ball is drawn. If the ball drawn at a particular trial is labeled B, the corresponding center is one where B is true; otherwise it is one where B is false.

Clauses (a) and (b) correspond to drawing from the urn with replacement, and drawing from the urn without replacement. Titelbaum's *Technicolor Beauty* example satisfies (a)—the procedure with the die and the colored paper is equivalent to drawing from an urn without replacement. The coin-toss variant on the *Technicolor Beauty* example satisfies (b)—the coin-toss procedure is equivalent to drawing from an urn without replacement.

At this point, we can exploit some useful probabilistic facts about expectation values. Recall the definition of the expected number of centers for a proposition A, from section 3.4.

$$C_A = \frac{\sum_{W \in A} Cr_u(W) N_W}{Cr_u(A)}$$

Where χ is any partition divides that the space of worlds into mutually exclusive, jointly exhaustive propositions, the Thirder Rule can be rewritten as follows in terms of expected numbers of centers.

$$Cr_{@}(A) = \frac{\sum_{X \in \chi} Cr_u(X \& A) Cx}{\sum_{X \in \chi} Cr_u(X) Cx}$$

When we enrich the original representation of the agent's belief worlds, her original belief worlds can be treated as members of a coarse-grained partition χ.

Using this fact, we can show that in examples satisfying either clause (a) or clause (b), enriching the agent's doxastically possible worlds has no effect on the Thirder Rule's prescriptions. First, consider examples that satisfy clause (a). In these examples, for each of the agent's original belief worlds W, the expected number of B centers at t_1 in W is δN_W in the enriched representation. (Intuitively, this is because at each original belief world, B has N_W chances to be true, each with probability δ.) Next, consider examples that satisfy (b). In such examples, at t_0 B is binomially distributed over the centers that will be doxastically possible at t_1. Thus, once again, within each original belief world W, the expected number of centers which are doxastically possible at t_1, and at which B is true, is δN_W.

Thus, the rewritten Thirder Rule tells us that in the enriched representation, where the W_s are worlds in the original representation and $Cr_{@}$ is the agent's actual t_1 credence function,

$$Cr_{@}(A) = \frac{\sum_{W} Cr_u(W \& A) \delta N_W}{\sum_{W} Cr_u(X) \delta N_W}$$

Or, to rephrase slightly,

$$Cr_{@}(A) = \frac{\sum_{W \in A} Cr_u(W) N_W}{\sum_{W^*} Cr_u(W^*) N_{W^*}}$$

This is just the original Thirder Rule. Different representation, same result.

Thus, the Thirder Rule boasts an advantage over the Halfer Rule. Enriching the representation of an agent's belief state spells trouble for halfers, but not thirders. Adding irrelevant *de se* information to a belief world seems to generate relevant *de dicto* information, since the more centers a world contains, the likelier it is to contain any given experience. Halfers have trouble accounting for this, since they hold that purely *de se* information is never relevant to *de dicto* matters. But thirders who treated *de se* information as relevant to *de dicto* matters all along will have no trouble. For thirders, the relevant *de dicto* information in the new representation is not really new information; it's just relevant *de se* information redescribed.

6. CONCLUSION

The *Sleeping Beauty* example shows that there is more to belief updating than conditionalizing on one's total evidence—at least if one's total evidence is taken to include *de se* information. Agents' belief states can be divided into centered and uncentered parts, and conditionalization can be understood as a constraint on the uncentered part. Halfers and thirders disagree about how the uncentered part of an agent's belief state should interact with the centered part to generate her credence function.

The answer to the debate between halfers and thirders turns on the answer to the debate between causal and evidential decision theorists: causal decision theorists should favor the Thirder Rule, while evidential decision theorists should favor the Halfer Rule. At first glance, this claim might seem bizarre. Why should one's commitments about *how to choose a course of action* influence one's beliefs about *the result of a coin toss*? Isn't this just a blatant conflation of pragmatic rationality with epistemic rationality?

It is indeed a conflation of pragmatic rationality with epistemic rationality, but I'm in good company. Justifications for adopting partial beliefs which correspond to the probability calculus almost invariably have a pragmatic component. Dutch books turn on the assumption that one should not have partial beliefs that expose one to a foreseeable sure loss—where 'sure loss' is cashed out in pragmatic terms. Representation theorems turn on qualitative requirements regarding agents' preferences, and requirements regarding preferences are surely pragmatic rather than epistemic. Scoring rules, while they may treat the goal of partial belief as non-pragmatic (See Joyce, 1998) still require commitments about the best means to attaining those goals. I haven't introduced new pragmatic commitments to the concept of partial belief—the pragmatic commitments were there all along.

When it comes to stability, the Thirder Rule outperforms the Halfer Rule. Since evidential decision theory entails a commitment to the Halfer Rule (at least for evidential decision theorists who countenance Dutch books and scoring rules), this is a point against evidential decision theory. Cases of irreducibly *de se* ignorance are rare, so perhaps the point is outweighed by evidential decision theory's good points. Still, the stability of the Thirder Rule is a hitherto unappreciated reason—even if a weak reason—for preferring causal decision theory over evidential decision theory.

REFERENCES

Arntzenius, Frank (2002). 'Reflections on Sleeping Beauty', *Analysis*, 62(2): 53–62.

Bostrom, Nick (2007). 'Sleeping Beauty and Self-location: A Hybrid Model', *Synthese*, 157: 59–78.

Bradley, Darren and Leitgeb, Hannes (2006). 'When Betting Odds and Credences Come Apart: More Worries for Dutch Book Arguments', *Analysis*, 66(2): 119–27.

Brier, Glenn (1950). 'Verification of Forecasts Expressed in Terms of Probability', *Monthly Weather Review*, 78(1): 1–3.

Draper, Kai and Pust, Joel (2008). 'Diachronic Dutch Books and Sleeping Beauty', *Synthese*, 164(2): 281–7.

Egan, Andy (2007). 'Some Counterexamples to Causal Decision Theory', *Philosophical Review*, 116(1): 93–114.

Elga, Adam (2000). 'Self-locating Belief and the Sleeping Beauty Problem', *Analysis*, 60(2): 143–7.

——(2004). 'Defeating Dr. Evil with Self-locating Belief', *Philosophy and Phenomenological Research*, 69(2): 383–96.

——(2007). 'Reflection and Disagreement', *Noûs*, 41: 479–502.

Hájek, Alan (2003). 'What Conditional Probability Could Not Be', *Synthese*, 137(3): 273–323.

Halpern, Joseph (2005). 'Sleeping Beauty Reconsidered: Conditionalization and Reflection in Asynchronous Systems', in Tamar Szabo Gendler (ed.), *Oxford Studies in Epistemology*, 1 (Oxford), 111–42.

Hitchcock, Chris (2004). 'Beauty and the Bets', *Synthese*, 139: 405–20.

Joyce, James (1998). 'A Non-pragmatic Vindication of Probabilism', *Philosophy of Science*, 65(4): 575–603.

Kierland, Brian and Monton, Bradley (2005). 'Minimizing Inaccuracy for Self-locating Beliefs', *PPR* 70(2): 384–95.

Lewis, David (1979). 'Attitudes *de dicto* and *de se*', *Philosophical Review*, 88(4): 513–43.

——(1986). 'A Subjectivist's Guide to Objective Chance', in *Philosophical Papers*, 2 (New York and Oxford).

——(1999). Why Conditionalize? In *Papers in Metaphysics and Epistemology* (Cambridge), 403–7.

——(2001). 'Sleeping Beauty: A Reply to Elga', *Analysis*, 61(3): 171–6.

McGee, Vann (2004). 'Learning the Impossible', in Ellery Eells and Brian Skyrms (eds), *Probability and Conditionals: Belief Revision and Rational Decision* (Cambridge), 179–99.

Meacham, Chris (2008). 'Sleeping Beauty and the Dynamics of *de se* Beliefs', *Synthese*, 148(2): 245–69.

Millikan, Ruth (1990). 'The Myth of the Essential Indexical', *Noûs*, 24(5): 723–34.

Perry, John (1971). 'The Problem of the Essential Indexical', *Noûs*, 13(1): 3–21.

—— (1974). 'Frege on Demonstratives', *Philosophical Review*, 86(4): 474–97.

Quine, W. V. (1969). 'Propositional Objects', in *Ontological Relativity and Other Essays* (New York).

Savage, Leonard (1971). 'Elicitation of Personal Probabilities and Expectations', *Journal of the American Statistical Association*, 66(336): 783–801.

Teller, Paul (1973). 'Conditionalization and Observation', *Synthese*, 26: 218–58.

Titelbaum, Mike (2008). 'The Relevance of Self-locating Beliefs', *Philosophical Review*, 117(4): 555–606.

Weatherson, Brian (2005). 'Should We Respond to Evil with Indifference?', *Philosophy and Phenomenological Research*, 70: 613–35.

2. Skeptical Success

Troy Cross

THE PROJECT

You intend to write a textbook on the theory of knowledge. Chapter 1 is slated to cover skepticism. You need some colorful skeptical scenarios, sure to vex your audience. Later, your anti-skeptical arguments will ease their pain, but for now, the goal is to inflict it.

"You think you know you have hands," you begin, "but maybe, *in fact*, you don't have any hands!"

You stop. Isn't that a skeptical scenario? It entails that you don't know something that you now take yourself to know, something you take yourself *obviously* to know. Yet somehow your hypothesis doesn't seem to live up to Descartes' dreaming argument (1985: 13) or his evil genius idea (1985: 15), Russell's five-minute hypothesis (2008: 104), Goodman's grue possibility (2006: 72), Goldman's phony barn country (1976: 772–3), or even the Wachowski brothers' Matrix movies (1999).[1]

You set yourself once more to the task. "You think you know you have hands," you begin, "but maybe you don't even *believe* you have hands! Belief is just as much a requirement for knowledge as truth, so if you don't even believe you have hands, you can't know you do."

This feels even worse. Your dream of Cartesian fame is fading rapidly.

One more try. "You think you know you have hands," you write, "and maybe you do have hands, but as it happens, you are correct just as a matter of luck! Lucky true belief is not knowledge, so you don't know you have hands."

A bit more subtle, a slight improvement over the others. Still, your scenario doesn't begin to approach the classics.

"What's missing?" you ask yourself. "It can't be that Descartes' scenarios are antecedently judged to be more *likely* than mine. Quite the opposite. I know probability theory. If I am a handless victim of an evil genius, then I am still handless. That's guaranteed. But I may be handless for reasons other

This paper grew out of comments I made on a paper by Jonathan Vogel and later a seminar I taught with Keith DeRose on knowledge and conditionals. Thanks to Jonathan, Keith, the participants in the seminar, and also the philosophy faculty at Yale, where an earlier version of the paper was discussed in a lunchtime colloquium. I'm also indebted to Peter Klein and Ernest Sosa, who started me thinking about these topics, and to Tamar Gendler, Roald Nashi, Amia Srinivasan, and Tamar Kreps for comments on earlier drafts.

[1] The phony barn scenario, though popularized by Alvin Goldman, is standardly attributed to Carl Ginet. In the paper I'll make use of the brain-in-a-vat scenario. Its origin is unknown, though like Keith Lehrer's (1971) 'Googol' example, it was probably a way of making the evil genius case compatible with materialism.