The Eternal Coin: a puzzle about self-locating conditional credence

CIAN DORR

It has been argued that self-locating credence is not reducible to credence about the world as a whole (Lewis 1979). It has also been argued that conditional credence is not reducible to unconditional credence (Hájek 2003). I will present a puzzle that arises from the interaction of the phenomena that motivate these two influential arguments.[1]

1 The question and two answers

The setup: The Eternal Coin has existed throughout an infinite past, and will continue to exist throughout an infinite future. It is fair, and it is tossed every day. It is causally isolated from you: you can never gain any evidence relevant to questions about how it lands.

Suppose you are ideally rational and that Cr is your conditional credence function, conditional on the setup. Let H be the centred proposition *the Coin lands Heads today*; let F be *the Coin will land Heads on every future day*; and let P be *the Coin has landed Heads on every past day*. What are $\mathrm{Cr}(H|F)$ and $\mathrm{Cr}(H|P)$?

On the assumption that these conditional credences are well-defined, there are two answers worth considering:

[1]Neither of the arguments is irresistible, of course. The phenomena that motivate taking self-locating credence as irreducible can perhaps be accommodated in a Fregean theory of propositions that allows for distinctive first-personal modes of presentation. And some of the phenomena that motivate taking conditional credence as irreducible can be accommodated in a theory that uses non-standard real numbers (including infinitesimals) to measure unconditional credence (McGee 1994). For my purposes it does not matter whether these reductions are successful. I will assume that unconditional and conditional credences are standard real numbers; if you favour the non-standard approach, you should translate everything I say by inserting 'the standard part of' in front of every expression that I treat as denoting a real number.

HALF $\mathrm{Cr}(H|F) = \mathrm{Cr}(H|P) = 1/2$

ONE $\mathrm{Cr}(H|F) = \mathrm{Cr}(H|P) = 1$

Prima facie, HALF looks obviously correct, and ONE looks like the result of a blatant fallacy. The obvious case for HALF can be spelt out as follows. According to the setup, the Eternal Coin is fair. The tosses of a fair coin are independent: the chance that the Coin lands Heads on any given day is 1/2, conditional on any consistent proposition about how the Coin lands on other days. Since you have no other relevant evidence, by the Principal Principle (Lewis 1980), your credences conditional on the setup should match what the setup says about the chances. So your credence that the Coin lands Heads today, conditional on the conjunction of the setup with any consistent hypothesis about what it does on other days, should be 1/2.

Of course, if you actually *learnt* that the Coin had landed Heads on every past toss, your unconditional credence in H should be much greater than 1/2, since that would be good evidence against the hypothesis that it is fair. But it would be an egregious mistake to think that this should affect your credence in *H conditional on the setup*, which entails that the Coin is fair.

However, there are arguments for ONE that do not commit this fallacy. I will present two of these arguments in the remainder of section 1; the remainder of the paper will be devoted to presenting and defending a third argument.

The argument from qualitative indiscernibility: The centred propositions F (*Heads from tomorrow on*) and HF (*Heads from today on*) describe “events of the same qualitative type”, which differ only in that one starts a day later. So you should regard them as equally likely. But for $\mathrm{Cr}(H|F)$ to be less than 1 would be for you to regard HF as less likely than F, and thus not to regard them as equally likely. So $\mathrm{Cr}(H|F) = 1$. For the same reason, you should regard P and HP as equally likely, and hence $\mathrm{Cr}(H|P) = 1$.[2]

[2]This argument is inspired by some moves in Williamson 2007, although the conclusion is certainly not one that Williamson would endorse. The step from ‘they describe events of the same qualitative type’ to ‘you should regard them as equally likely’ corresponds to a key step in Williamson’s argument. (Indeed, this inference is on a firmer footing in the case of the Eternal Coin than in Williamson’s case, which involves an sequence of coin-tosses with a first element

The argument from qualitative indiscernibility is not compelling. In a framework where conditional probabilities are being taken as primitive, it is natural to distinguish weak and strong senses of 'regard as equally likely'. To regard two hypotheses as equally likely in the weak sense is to assign them the same unconditional credence; to regard them as equally likely in the strong sense is to assign them the same credence conditional on their disjunction. It is plausible that you should regard F and HF as equally likely in the weak sense: your unconditional credence in each of them should be zero. But since we allow nontrivial credences conditional on hypotheses with unconditional credence zero, this is consistent with the claim that $\mathrm{Cr}(H|F) = 1/2$. On the other hand, the argument that you should regard HF and F as equally likely in the strong sense is unconvincing. Even though the hypotheses are (in a natural sense) qualitatively indiscernible, they clearly bear different qualitative *relations* to the disjunction $F \vee HF$: F is entailed by (indeed, equivalent to) the disjunction and HF is not. So in assigning F and HF different credences conditional on $F \vee HF$, one need not break any qualitative symmetries of the setup.[3]

Our second argument for ONE is harder to dismiss.

The argument from self-locating indifference: Consider a simplified variant of the much-discussed story of Sleeping Beauty. For reasons we need not go into, you suspect that on Monday night you will be given a drug that will erase all your memories from Monday. When you wake up and the last thing you remember is Sunday, you should be uncertain whether it is Monday or Tuesday. In fact, assuming that you have no other relevant evidence, it seems obvious that conditional on the hypothesis that you are given the drug, your credence that it is Monday should be *equal* to

but no last element—in Williamson's case, the event of the coin's landing Heads every day differs from the event of the coin's landing Heads on the second and every subsequent day in extrinsic, but nevertheless qualitative, respects, and it is not clear why this difference should be irrelevant to your credences.) The idea that regarding A and AB as equally likely requires $\mathrm{Cr}(B|A)$ to equal 1 if it is defined at all, on the other hand, may be suggested by a brief discussion of preference in Williamson's penultimate paragraph, but is not central to the goal of his paper.

[3] There may, however, be a good reason here to reject the reduction of conditional probability to non-standard-valued unconditional probabilities: according to that theory, if $\mathrm{Cr}(H|F) = 1/2$, your unconditional credence in HF must be an infinitesimal number that is half of the infinitesimal number that is your unconditional credence in F, so your unconditional credences do, arguably, break the qualitative symmetries of the setup. The idea that your credences in hypotheses like F and HF should be infinitesimal but not zero is the central target of Williamson 2007.

your credence that it is Tuesday. This seems to have been universally accepted in the literature on the Sleeping Beauty puzzle (see, e.g., Elga 2000, Lewis 2001, Dorr 2002)—the controversy there is about how confident you should be in the hypothesis that you are given the drug, not about your credences conditional on that hypothesis. This intuition is especially compelling when we stipulate that the drug will make your Tuesday morning experiences exactly match your Monday morning experiences in all respects. But the duplication of experience is not really crucial—it should not matter if we allow some realistic differences between the details of your Monday evidence and the details of your Tuesday evidence. What matters is just that your evidence does nothing to discriminate between the hypothesis that it is Monday and you will receive the drug and on the hypothesis that it is Tuesday and you have received the drug.

This intuition is an an application of an appealing principle of indifference for self-locating belief about the time. Roughly speaking: when your evidence does not discriminate between different hypotheses about what time it is, you should divide your credence equally between them. Less roughly: Take any jointly inconsistent centred propositions A and B that agree about what the world is like as a whole and who you are, and disagree only about when it is. Suppose that if either A and B is ever true of you, each must each be true of you for the same length of time—e.g. one day. And suppose that your evidence does nothing to discriminate between A and B, in the sense that *a priori* you should regard it as no more or less likely given A than given B. Then you should regard A and B as equally likely, in the strong sense: conditional on $A \vee B$, your credences in A and in B should be equal.[4]

This principle of indifference entails ONE. I will show how the argument goes in the case of the claim that $\mathrm{Cr}(H|F) = 1$; the argument that $\mathrm{Cr}(H|P) = 1$ is exactly analogous.

Let K_0 be the centred proposition *today is the last Tails day preceding an infinite final run of*

[4]A similar principle of self-locating indifference is stated and defended in Elga 2004. Elga's principle is weaker than mine in two respects: it only applies to centred propositions that are true relative to only one person at only one instant of time in one possible world, and it only applies to pairs of such propositions that are "subjectively indistinguishable". It is stronger than mine in one respect: the pairs of centred propositions to which it applies can differ not only as regards the time, but also as regards the identity of the person at the centre.

Heads. When $0 < i < \infty$, let K_i be the centred proposition *today is the ith day of an infinite final run of Heads*. Finally, let K_∞ be the proposition *the Coin lands Heads every day, past present and future*. Given that it is part of the setup that you never gain evidence relevant to questions about how the Coin lands, your evidence does nothing to discriminate between K_i and K_j (when $0 \leq i < j < \infty$). So, according to the indifference principle, you should regard K_i and K_j as equally likely, in the strong sense. Since they are jointly inconsistent, this means that $\mathrm{Cr}(K_i|K_i \vee K_j) = 1/2$. That is: on the assumption that there is an infinite final sequence of Heads days and that tomorrow is one of them, you should think it no more likely that tomorrow is the $(i + 1)$st day than that tomorrow is the $(j + 1)$st day. Since the K_i are mutually inconsistent, it follows from this that $\mathrm{Cr}(K_0|K_0 \vee K_1 \vee \ldots \vee K_n) = 1/n$, for each n. Hence your credence in K_0 conditional on the infinite disjunction $K_0 \vee K_1 \vee \ldots$ must be less than $1/n$ for every n, which means it must be 0. (Remember that we are taking conditional credences to be standard real numbers.) Since each K_i except for K_0 entails H, it follows that $\mathrm{Cr}(H|K_0 \vee K_1 \vee \ldots) = 1$. Also, since K_∞ entails H, $\mathrm{Cr}(H|K_\infty) = 1$. But F is equivalent to the disjunction $K_0 \vee K_1 \vee K_2 \vee \ldots \vee K_\infty$; so $\mathrm{Cr}(H|F)$ must also be 1.

By contrast, upholding HALF requires a striking failure of the self-locating indifference principle. According to Half, you must regard HF and the conjunction of not-H with F as equally likely, in the strong sense. But the former is equivalent to the disjunction $K_1 \vee K_2 \vee \ldots \vee K_\infty$, while the latter is equivalent to K_0. According to HALF, then, you must regard K_0 as being at least as likely as the disjunction of all the remaining K_i. On the assumption that tomorrow lies somewhere in a final all-Heads sequence, you think it at least as likely as not that it is the very first day. Assuming that you extend HALF in the natural way to centred propositions about other days, your conditional credences will be 1/2 that it is the first day, 1/4 that it is the second day, 1/8 that it is the third day, and so on. This is a very dramatic departure from indifference.

2 The argument from evolving credences

The premise of the argument from self-locating indifference is quite intuitive, but it is not clear whether there is anything of a more systematic nature to be said in support of it. And as a general matter, indifference principles have a well-deserved bad name. By contrast, it is clear that the Principal Principle or something very like it plays a central role in all of our reasoning concerning chance, and that such reasoning is of great importance throughout the sciences. In the conflict between self-locating indifference (favouring ONE) and the Principal Principle (favouring HALF), conservativism legitimately favours the latter.

This section will bolster the case for ONE by presenting a third argument, which does not rely on any indifference-style assumptions. It has three two-part premises. For the purposes of stating these premises, let Cr_+ be the conditional credence function you will have tomorrow, conditional on the setup.

UPDATING a. $Cr_+(HF|P \vee HF) = Cr(F|HP \vee F)$

b. $Cr_+(P|P \vee HF) = Cr(HP|HP \vee F)$[5]

Preliminary defence: You are not going to gain or lose any evidence relevant to questions about the Coin between today and tomorrow. So if you are rational, you will not change your mind about the Coin. But of course we must take account of the fact that the day you think about as 'today' today is the one you will think about as 'yesterday' tomorrow. In the intended sense of 'changing your mind', you can count as not changing your mind even if your credence in the centred proposition *it is raining today* is high today and low tomorrow: what matters is that your credence today in *it is raining today* should match your credence tomorrow in *it was raining yesterday*. Similarly, then, if you do not change your mind about the Coin, your credential attitudes towards F and HP today must be the same as your credeential attitudes towards HF and P tomorrow.

CONSTANCY a. $Cr_+(HF|P \vee HF) = Cr(HF|P \vee HF)$

b. $Cr_+(P|P \vee HF) = Cr(P|P \vee HF)$[6]

Preliminary defence: Your total evidence (both centred and uncentred) does not change in any relevant way between today and tomorrow. So your conditional credences in centred propositions

[5]Note that (a) and (b) are equivalent, given the very plausible assumption that $Cr(HPF|HP \vee F) = Cr_+(HPF|P \vee HF) = 0$.

[6](a) and (b) here are also equivalent if $Cr(HPF|HP \vee F) = Cr_+(HPF|P \vee HF) = 0$.

about the Coin should not change either. By contrast, in run-of-the mill cases where the mere passage of time makes a difference to your credences in centred propositions, there is a corresponding change in your centred evidence. For example, in an ordinary case where the passage of time diminishes your credence in the centred proposition *it is raining*, your evidence today may include the centred proposition *it is raining right now*, or perhaps *I seem to see that it is raining right now*, while your evidence tomorrow instead includes the centred proposition *it was raining yesterday*, or perhaps *I seem to remember its raining yesterday*. Nothing like this happens with the Coin.

POSITIVITY a. $\mathrm{Cr}(P|P \vee F) > 0$

b. $\mathrm{Cr}(F|P \vee F) > 0$

Preliminary defence: For POSITIVITY to fail would be for you to regard one of F and P as infinitely more likely than the other. Such a bias seems odd, given the symmetry of the setup. Compare the present case to one where infinitely many coins are tossed simultaneously, in an infinite row of houses running East-West. In that case, it would be crazy to regard *Heads in every house to the East of me* as infinitely more or less likely than *Heads in every house to the West of me*. Why should it be any different when the distribution is temporal rather than spatial?

Besides these three premises, the only other ingredients we need for the argument are (MA) the multiplicative axiom for conditional probability, according to which $\mathrm{Cr}(A|C) = \mathrm{Cr}(A|B)\,\mathrm{Cr}(B|C)$ whenever AC entails B and B entails C, and the principle (UI) that conditional probabilities always belong to the unit interval $[0, 1]$.

Here is the argument. First, combining UPDATING and CONSTANCY yields

(1) a. $\mathrm{Cr}(F|HP \vee F) = \mathrm{Cr}(HF|P \vee HF)$

b. $\mathrm{Cr}(HP|HP \vee F) = \mathrm{Cr}(P|P \vee HF)$

Next, using POSITIVITY and (1), we derive formulae for $\mathrm{Cr}(H|F)$ and $\mathrm{Cr}(H|P)$:

(2) a. $$\begin{aligned}\mathrm{Cr}(H|F) &= \frac{\mathrm{Cr}(HF|P \vee F)}{\mathrm{Cr}(F|P \vee F)} && \text{by MA and POSITIVITY (a)}\\ &= \frac{\mathrm{Cr}(HF|P \vee HF)\,\mathrm{Cr}(P \vee HF|P \vee F)}{\mathrm{Cr}(F|HP \vee F)\,\mathrm{Cr}(HP \vee F|P \vee F)} && \text{by MA}\\ &= \frac{\mathrm{Cr}(P \vee HF|P \vee F)}{\mathrm{Cr}(HP \vee F|P \vee F)} && \text{by (1a).}\end{aligned}$$

b. $$\begin{aligned}\mathrm{Cr}(H|P) &= \frac{\mathrm{Cr}(HP|P \vee F)}{\mathrm{Cr}(P|P \vee F)} && \text{by MA and POSITIVITY (b)}\\ &= \frac{\mathrm{Cr}(HP|HP \vee F)\,\mathrm{Cr}(HP \vee F|P \vee F)}{\mathrm{Cr}(P|P \vee HF)\,\mathrm{Cr}(P \vee HF|P \vee F)} && \text{by MA}\\ &= \frac{\mathrm{Cr}(HP \vee F|P \vee F)}{\mathrm{Cr}(P \vee HF|P \vee F)} && \text{by (1b).}\end{aligned}$$

(2a) and (2b) jointly entail that

(3) $$\mathrm{Cr}(H|F) = \frac{1}{\mathrm{Cr}(H|P)}$$

But given UI, (3) can only be true if $\mathrm{Cr}(H|F) = \mathrm{Cr}(H|P) = 1$.

3 Further defence of the premises

Since the argument from evolving credences has three premises, there are three minimal ways for defenders of HALF to resist it. This section will address these; section 4 will consider some more radical responses.

Keep UPDATING *and* CONSTANCY*, reject* POSITIVITY. To uphold HALF in this way, we would have to claim that it is rationally *forbidden* for both $\mathrm{Cr}(P|P \vee F)$ and $\mathrm{Cr}(F|P \vee F)$ to be positive. Otherwise, the argument of section 2 would still show that is rationally *permissible* for $\mathrm{Cr}(H|F)$ and $\mathrm{Cr}(H|P)$ to be 1. But it is hard to see where such a rational requirement could come from. Is there some specific one of P and F that you are rationally required to regard as infinitely more likely than the other? If so, which, and why? On the other hand, if rationality permits regarding *either* hypothesis as infinitely more likely, it seems odd that it should forbid the less extreme intermediate attitudes in which one regards one as only finitely more likely than the other.

Even if there were some such prohibition, HALF would not be safe. For we can consider slight variations of the setup that add lots of unrelated chancy propositions with unknown truth-values. Even if, say, $\mathrm{Cr}(F|P \vee F) = 0$, if the background space of uncertain propositions is rich, it will plausibly contain some Q that you regard as just unlikely enough that $\mathrm{Cr}(PQ|PQ \vee F)$ and $\mathrm{Cr}(F|PQ \vee F)$ are both positive. But then, by replacing P with PQ throughout the argument, we will still be able to show that $\mathrm{Cr}(H|F) = 1$. Similarly, if we can find some Q such that $\mathrm{Cr}(P|P \vee (F \vee Q))$ and $\mathrm{Cr}(F \vee Q|P \vee (F \vee Q))$ are both positive, we can still show that $\mathrm{Cr}(H|P) = 1$, by substituting $F \vee Q$ for F throughout the argument.

Keep UPDATING *and* POSITIVITY, *reject* CONSTANCY. This is a view on which your centred conditional credences about the Coin change from one day to the next, despite the fact that there is no relevant change in your centred evidence. As time goes on, you regard HF as more and more likely relative to P. Given HALF, we can quantify this change. If the ratio of your credences in HF and P, conditional on their disjunction, is x today, tomorrow it will be $4x$, since HALF requires you to regard F as twice as likely as HF and HP as half as likely as P, and UPDATING requires your attitudes toward HF and P tomorrow to match your attitudes toward F and HP today.[7] By similar reasoning, the ratio the day after tomorrow will be $16x$. j days from now, it will be $4^j x$. j days ago, it was $4^{-j} x$. You spend almost all of your life regarding one of P and HF as vastly more likely than the other.

How could you get your conditional credences to evolve like this? In settling on a credence today, you would have to consult your memories of your credences on previous days, and fix your credence today as a function of these. This procedure is odd: it seems to involve your treating facts about your own past mental states as if they were evidence bearing on hypothesis about how the Coin lands. To bring out the oddness of this, imagine that you wake up each morning with a general knowledge of the setup but no particular memories of your past; at noon, your amnesia dissipates. The ratio of your credences in HF and P, conditional on their disjunction,

[7] *Proof:* By HALF and MA, (A) $\mathrm{Cr}(HF|P \vee F) = \mathrm{Cr}(H|F)\,\mathrm{Cr}(F|P \vee F) = 1/2\,\mathrm{Cr}(F|P \vee F)$, and (B) $\mathrm{Cr}(HP|P \vee F) = \mathrm{Cr}(H|P)\,\mathrm{Cr}(P|P \vee F) = 1/2\,\mathrm{Cr}(P|P \vee F)$. Then $\mathrm{Cr}_+(HF|P \vee HF)/\,\mathrm{Cr}_+(P|P \vee HF)$ $= \mathrm{Cr}(F|HP \vee F)/\,\mathrm{Cr}(HP|HP \vee F)$ (by UPDATING) $= (\mathrm{Cr}(F|P \vee F)/\,\mathrm{Cr}(HP \vee F|P \vee F))/(\mathrm{Cr}(HP|P \vee F)/\,\mathrm{Cr}(HP \vee F|P \vee F))$ (by MA and the fact that $\mathrm{Cr}(HP \vee F|P \vee F)$ is positive, which follows from POSITIVITY) $= \mathrm{Cr}(F|P \vee F)/\,\mathrm{Cr}(HP|P \vee F) = 2\,\mathrm{Cr}(HF|P \vee F)/(1/2)\,\mathrm{Cr}(P|P \vee F)$ (by A and B) $= 4\,\mathrm{Cr}(HF|P \vee F)/\,\mathrm{Cr}(P|P \vee F) = 4\,\mathrm{Cr}(HF|P \vee HF)\,\mathrm{Cr}(P \vee HF|P \vee F)/\,\mathrm{Cr}(P|P \vee HF)\,\mathrm{Cr}(P \vee HF|P \vee F)$ (by MA) $= 4\,\mathrm{Cr}(HF|P \vee HF)/\,\mathrm{Cr}(P|P \vee HF)$.

will not vary systematically from one morning to the next. Suppose that this morning the ratio was 1 : 1. At noon, you remember that yesterday afternoon the ratio was 8 : 1. Could you now be rationally obliged to set the ratio to 32 : 1, so as to stay in harmony with your earlier self? It seems doubtful. When your memories come back, you should update by conditionalising on what you learn.[8] But during the morning, your credences about the Coin should have been independent of your credences about your own past mental states. Since you knew that the beliefs you had yesterday were not based on any special evidence or insight, you had no reason to regard them as correlated in any particular way with facts about the Coin. Ideally rational believers have no need to accord special evidential weight to facts about their past beliefs in order to achieve the kind of diachronic harmony that motivates UPDATING. They achieve it automatically, by always using the same method to base their credences on their changing total evidence.

Keep CONSTANCY *and* POSITIVITY*, reject* UPDATING. By CONSTANCY, the ratio of $\mathrm{Cr}(HF|P \vee HF)$ to $\mathrm{Cr}(P|P \vee HF)$ is the same each day: suppose for concreteness that it is 1 : 1. Then, given the obvious generalisation of HALF to centred propositions about what the Coin does on other days, the ratio of your credences in *Heads every day up to today and for the next nine days* and *Heads every day starting ten days from now*, conditional on their disjunction, is $2^{-10} : 2^{10}$. You are almost certain that if one of these hypotheses is true, it is the second. Similarly, you are almost certain that if one of *heads every day up to 11 days ago* and *heads every day starting ten days ago* is true, it is the first. Since your attitudes to centred propositions about the Coin do not change from day to day, all this will be the same twenty days hence. So you can make the following speech on October 1st:

> If October 11th either has the feature of being preceded only by Heads days, or the feature of being a Heads day that is followed only by Heads days, then almost certainly it has the latter. But by October 21st I will have changed my mind about this: I will then be almost certain that if it has either feature it has the former. So almost certainly, if October 11th has either feature, I will be wrong on October 21st about which one

[8] Cf. Dorr 2002.

it has if it has either. And I will not have become irrational, or forgotten some of my current evidence, or lost track of the time, or anything like that. The method I will be using then to form my almost certainly inaccurate conditional belief is the exact same method I am using today to form my almost certainly accurate conditional belief.

I do not think a rational person could acquiesce in this way to a (conditional) prediction of future (conditional) error. It is not obvious how to formulate a general principle that captures what is wrong with such self-mistrust: the best known attempt, van Fraassen's Reflection Principle, is subject to decisive objections.[9] I will not attempt here to derive UPDATING from any such general principle: it is clearer that the above speech manifests some kind of irrationality than what, exactly, that irrationality consists in.

4 Are the relevant conditional credences well-defined?

One could take the puzzle as a *reductio* of the assumption that $\mathrm{Cr}(H|P)$ and $\mathrm{Cr}(H|F)$ are well-defined at all. This would presumably require denying that one can ever have well-defined credences conditional on hypotheses with unconditional credence zero, since examples very close to these are the central exhibits in the case for allowing such conditional credences sometimes to be well-defined.[10] If we did deny this, we would be under pressure also to deny the possibility of well-defined *chances* conditional on hypotheses whose unconditional *chance* is zero. For if one could rationally assign positive credence to the hypothesis (X) that $\mathrm{chance}(A|B) = x$ and $\mathrm{chance}(B) = 0$, then *prima facie*, one's credence in B given X should be 0 and one's credence in A given BX should be x. Allowing primitive conditional chances while rejecting primitive conditional credences would require a pointless circumscribing of the Principal Principle.

It would be an overreaction to give up on primitive conditional credences and chances altogether: they are useful in many ways. For example, we need nontrivial chances conditional on

[9] See van Fraassen 1984, Elga 2007, Briggs 2009.

[10] A close relative of this view would claim that $\mathrm{Cr}(A|B)$ is trivially equal to 1 whenever $\mathrm{Cr}(B) = 0$. Proponents of this view need have no objection to the argument from evolving credences, although they will regard it as needlessly complex.

propositions whose unconditional chance is zero in order to straightforwardly express how the chances at earlier times constrain the chances at later times: the chance function at t_2 derives from the chance function at t_1 by conditionalising on the complete truth about history between t_1 and t_2, whose chance at t_1 may well have been 0.

Defenders of HALF obviously cannot reject conditionalisation on credence-zero hypotheses in general. But they might still want to deny that the specific conditional credences that figure in the argument from evolving credences—things like $\mathrm{Cr}(P|P \vee F)$ and $\mathrm{Cr}(F|P \vee F)$—are well-defined. It would be implausible to reject the well-definedness of *all* credences conditional on $P \vee F$—surely, given that $\mathrm{Cr}(H|F)$ and $\mathrm{Cr}(H|P)$ both equal $1/2$, $\mathrm{Cr}(H|P \vee F)$ must also be $1/2$. But perhaps there is something special about P and F which prevents $\mathrm{Cr}(P|P \vee F)$ and $\mathrm{Cr}(F|P \vee F)$ from being well-defined.

A natural mathematical setting for such localized failures of well-definedness is the theory of "unsharp credences", in which a person's belief state is represented by a *set* of (conditional) probability functions—a 'representor'—rather than by a single (conditional) probability function.[11] In the special case where $\mathrm{C}(A|B) = x$ for each C in your representor, we can say that your conditional credence in A given B is x; but when the functions in your representor take different values given A and B as arguments, there is no number that is your conditional credence in A given B. Those who endorse this framework can resist the argument from evolving credences by claiming that if you are rational, the functions in your representor must differ in the conditional probabilities they assign to pairs of centred propositions like P and $P \vee F$.

On this view, UPDATING, CONSTANCY, and POSITIVITY must all be rejected as involving ill-defined quantities. But interestingly, defenders of HALF can consistently accept modified versions of all three premises, which preserve much of their spirit in an unsharp setting. (Here **R** is

[11] See Levi 1974, Joyce 2005.

your representor today, and $\mathbf{R}_+$ is your representor tomorow.)

UNSHARP UPDATING

a. $$\{C(HF|P \vee HF) : C \in \mathbf{R}_+\} = \{C(F|HP \vee F) : C \in \mathbf{R}\}$$

b. $$\{C(P|P \vee HF) : C \in \mathbf{R}_+\} = \{C(HP|HP \vee F) : C \in \mathbf{R}\}$$

UNSHARP CONSTANCY

a. $$\{C(HF|P \vee HF) : C \in \mathbf{R}_+\} = \{C(HF|P \vee HF) : C \in \mathbf{R}\}$$

b. $$\{C(P|P \vee HF) : C \in \mathbf{R}_+\} = \{C(P|P \vee HF) : C \in \mathbf{R}\}$$

UNSHARP POSITIVITY

a. $$0 \notin \{C(P|P \vee F) : C \in \mathbf{R}\}$$

b. $$0 \notin \{C(F|P \vee F) : C \in \mathbf{R}\}$$

The key to the consistency of this package is the fact that even if none of the particular conditional probability functions that make up your representor is invariant under the function that "evolves centred propositions forward by a day", taking F to HF and HP to P, your whole representor can still be invariant under the action of this function.

Together with HALF—which in this context means that $C(H|P) = C(H|F) = 1/2$ for each $C \in \mathbf{R}$—these modified versions of our three premises require the members of $\mathbf{R}$ to have *arbitrarily strong* disagreements about the relative probability of P and HF. Suppose that for some $C \in \mathbf{R}$, $C(HF|P \vee HF)/C(P|P \vee HF) = x$. Then $C(F|HP \vee F)/C(HP|HP \vee F) = 4x$ (see note 7). But then, by UNSHARP UPDATING and UNSHARP CONSTANCY, $\mathbf{R}$ must contain some C^* such that $C^*(HF|P \vee HF)/C^*(P|P \vee HF) = 4x$. Since this argument can be iterated in both directions, the set $\{C(HF|P \vee HF)/C(P|P \vee HF) : C \in \mathbf{R}\}$ has no finite positive upper or lower bound.

My main worries about this response are worries about the unsharp credence framework itself. In my view, there is no adequate account of the way unsharp credences should be manifested in decision-making. As Adam Elga (MS) has recently compellingly argued, the only viable strategies which would allow for someone with an unsharp credential state to maintain a reasonable pattern of behavioural dispositions over time involve, in effect, choosing a particular member of the representor as the one that will guide their actions. (The choice might be made at the outset, or might be made by means of a gradual process of narrowing down over time; the upshot is much the same.) And even though crude behaviourism must be rejected, I think that if this is all we have to say about the decision theory, we lack an acceptable account of *what it is* to be in a given unsharp credential state—we cannot explain what would constitute the difference between someone in a sharp credential state given by a certain conditional probability function, and someone in an unsharp credential state containing that probability function, who had chosen is as the guide to their actions. Unsharp credential states seem to have simply been postulated as states that get us out of tricky epistemological dilemmas, without an adequate theory of their underlying nature. It is rather as if some ethicist were to respond to some tricky ethical dilemma—say, whether you should join the Resistance or take care of your ailing mother—by simply postulating a new kind of action that is stipulated to be a special new kind of combination of joining the Resistance and taking care of your mother which lacks the objectionable features of obvious compromises (like doing both on a part-time basis or letting the outcome be determined by the roll of a dice). It would be epistemologically very convenient if there was a psychological state we could rationally be in in which we neither regarded P as less likely than HF, regarded HF as less likely than P, nor regarded them as equally likely. But we should be wary of positing psychological states for the sake of epistemological convenience.[12]

This is not the right place for a full-dress critique of unsharp credences. Let me merely remind friends of HALF that even if they invoke the machinery of unsharp credence to escape the argument from evolving credences, they will still need something to say about the argument from self-locating indifference from section 1. They will have to explain away the appeal of the indifference principle that drives that argument, preferably in a way that explains why it gives correct answers in ordinary cases. There is no obvious way for unsharp credences to help with this task, since HALF requires your relative credences in centred propositions of the form *today is the ith day of a final infinite run of Heads* to depart from indifference in a very specific, and perfectly

[12]For another recent argument against unsharp credence, see White 2010 and the response in Sturgeon 2010.

sharp, way.

5 Non-repeating sequences of outcomes

Consider any function σ from the integers to $\{0, 1\}$. Let H_σ, P_σ and F_σ stand for the following centred propositions:

> H_σ: *The Coin lands Heads today iff* $\sigma(0) = 1$
>
> P_σ: *For each* $n > 0$*, the Coin landed Heads n days ago iff* $\sigma(-n) = 1$
>
> F_σ: *For each* $n > 0$*, the Coin will land Heads n days from now iff* $\sigma(n) = 1$.

Using this notation, we can state analogues of the premises of the argument from evolving credences—call them σ-UPDATING, σ-CONSTANCY and σ-POSITIVITY—which jointly entail that $\mathrm{Cr}(H_\sigma|P_\sigma) = \mathrm{Cr}(H_\sigma|F_\sigma) = 1$. If this were true for every σ, it would be a disaster for the friend of non-trivial conditionalisation on credence-zero propositions. For, by considering any σ and σ' which differ only at 0, we could deduce that even contradictions have credence 1 conditional on any P_σ or F_σ.

Fortunately, the considerations that support UPDATING do not extend to σ-UPDATING (for arbitrary σ). In general, there is no special reason to think that my attitudes towards $H_\sigma F_\sigma$ and P_σ tomorrow should match my attitudes towards F_σ and $H_\sigma P_\sigma$ today. Whereas HF is the result of "evolving F forward by one day", unless σ is trivial, evolving F_σ forward by one day does not yield $H_\sigma F_\sigma$. Rather, it yields distinct centred proposition $H_{\sigma_+} F_{\sigma_+}$, where $\sigma_+(n) := \sigma(n+1)$. So the considerations of diachronic harmony that support UPDATING do not support σ-UPDATING, but rather (4):

(4) (i) $\mathrm{Cr}_+(H_{\sigma_+} F_{\sigma_+} | P_{\sigma_+} \vee H_{\sigma_+} F_{\sigma_+}) = \mathrm{Cr}(F_\sigma | H_\sigma P_\sigma \vee F_\sigma)$

(ii) $\mathrm{Cr}_+(P_{\sigma_+} | P_{\sigma_+} \vee H_{\sigma_+} F_{\sigma_+}) = \mathrm{Cr}(H_\sigma P_\sigma | H_\sigma P_\sigma \vee F_\sigma)$

When σ is a periodic sequence, we can still use (4), in conjunction with σ-CONSTANCY and σ-POSITIVITY, to argue that $\mathrm{Cr}(H_\sigma|P_\sigma) = \mathrm{Cr}(H_\sigma|F_\sigma) = 1$. But for a *typical* σ, with no interesting periodic behaviour, there is nothing to defeat the presumption that $\mathrm{Cr}(H_\sigma|P_\sigma) = \mathrm{Cr}(H_\sigma|F_\sigma) = 1/2$.

In fact, there is good reason to think that you should have credence 1 that the σ which corresponds to the actual sequence of outcomes is one of those for which $\mathrm{Cr}(H_\sigma|P_\sigma) = \mathrm{Cr}(H_\sigma|F_\sigma) = 1/2$.[13]

Define $r(\sigma) := \mathrm{Cr}(F_\sigma|P_\sigma \vee F_\sigma)/\mathrm{Cr}(P_\sigma|P_\sigma \vee F_\sigma)$. If $r(\sigma)$ is high, you regard σ as more likely to be right about the future behaviour of the coin than about the past behaviour of the coin; if $r(\sigma)$ is low, the reverse is true. I know of no non-arbitrary, general rule which could be used to settle the value of $r(\sigma)$ for a typical σ without any symmetry or periodic behaviour in either direction. We cannot, for example, just set $r(\sigma) = 1$ for every σ: for by (4) and the argument of note 7, $r(\sigma_+) = 4r(\sigma)$, assuming that $\mathrm{Cr}(H_\sigma|P_\sigma) = \mathrm{Cr}(H_\sigma|F_\sigma) = \mathrm{Cr}(H_{\sigma_+}|P_{\sigma_+}) = \mathrm{Cr}(H_{\sigma_+}|F_{\sigma_+}) = 1/2$. This means we can make $r(\sigma)$ as small or as big as we like just by shifting σ along to the left or to the right. And for a typical σ, there is no non-arbitrary principle which could tell us how far to the left or right we need to shift σ to get a "balanced" sequence with an r-value close to one.

This kind of arbitrariness elicits different reactions. Some will be happy with an appeal to epistemic permissiveness: ideal rationality requires one to settle on a particular r, but any r that meets certain minimal constraints is as good as any other. Friends of unsharp credences will probably think that we should avoid this arbitrary choice by having a large representor that contains a conditional probability function for each r meeting the minimal constraints. The biggest challenge is for those (e.g. White 2005, 2010), who reject both permissiveness and unsharpness. They will have to

[13]The reasoning I have in mind depends on two plausible premises. (i) Your credences are conglomerable in the partitions $\{P_\sigma\}$ and $\{F_\sigma\}$ (see Kadane et al. 1986; Arntzenius et al. 2004). (ii) Whenever a centred proposition B is entirely about what the coin does in the past and future, and $\mathrm{Cr}(B)$ is positive, $\mathrm{Cr}(H|B) = 1/2$.

Let P_0 be the disjunction of all P_σ such that $\mathrm{Cr}(H|P_\sigma) = 0$ and P_1 the disjunction of all P_σ such that $\mathrm{Cr}(H|P_\sigma) = 1$. By (i), $\mathrm{Cr}(H|P_0) = 0$ and $\mathrm{Cr}(H|P_1) = 1$. So by (ii), $\mathrm{Cr}(P_0) = \mathrm{Cr}(P_1) = 0$. Since P_σ is a disjunct of either P_0 or P_1 whenever $\mathrm{Cr}(H_\sigma|P_\sigma) = 1$, your credence in the disjunction of all such P_σ must be 0. Assuming that $\mathrm{Cr}(H_\sigma|P_\sigma)$ is always either 1/2 or 1, it follows that you have credence 1 in the disjunction of all P_σ for which $\mathrm{Cr}(H|P_\sigma) = 1/2$. (And by parallel reasoning, you have credence 1 in the disjunction of all F_σ for which $\mathrm{Cr}(H|F_\sigma) = 1/2$.)

Even without the assumption that $\mathrm{Cr}(H_\sigma|P_\sigma)$ is always either 1/2 or 1, (i) and (ii) entail that whenever you have positive credence in the disjunction of all P_σ such that $\mathrm{Cr}(H_\sigma|P_\sigma) \in [a, b]$, $1/2 \in [a, b]$. This entails that whenever $a < 1/2 < b$, you have credence 1 in the disjunction of all P_σ for which $\mathrm{Cr}(H_\sigma|P_\sigma) \in (a, b)$. In the absence of countable additivity—which evidently must be rejected by anyone who accepts ONE—this does not require you to have credence 1 in the disjunction of the P_σ for which $\mathrm{Cr}(H_\sigma|P_\sigma)$ is *exactly* 1/2. But it comes very close.

say that there is a particular *r* that is generated by the one true ideally rational conditional credence function; the challenge is to say what it is. One could refuse this challenge on the grounds that we just cannot know what the privileged *r* is, or on the grounds that there is no fact of the matter about what it is; but each of these options has its own costs.

6 Clarifying the role of self-locating uncertainty

Consider worlds that contain, as well as the Eternal Coin, a gong that rings just once. There are at least two different ways to set up a conditional chance function over such worlds, each of which has a claim to represent the coin as fair, the gong as equally likely to ring on any given day, and the two as independent.[14] Roughly speaking, we could either start with a conditional probability function over propositions about how the coin lands each day and then fine-grain its domain to include propositions about the gong, or else we could start with a conditional probability function over propositions about when the gong rings and then fine-grain its domain to include propositions about the coin. On a "coin first" chance function, conditional on there being an initial or final infinite sequence of Heads outcomes, the gong is equally likely to ring on the *i*th day of the sequence as on the *j*th day. For reasons that emerged in our discussion of the argument from self-locating indifference, this means that the chance that the coin lands Heads on the day the gong rings, conditional on its landing Heads on every subsequent or preceding day, is 1. On a "gong first" chance function, by contrast, propositions about what the coin does on the day the gong rings are independent of propositions about what it does on other days, so this conditional chance is 1/2.

We could try arguing that only one of these two kinds of conditional probability function represents the coin and gong as *genuinely* independent. But not much hangs on this.Provided that you are a realist about chance (or can speak as one for present purposes), and that you allow non-trivial conditional conditionalization on chance-zero propositions, you should think that the hypothesis that the actual chance function is "coin first" and the hypothesis that it is "gong first" are coherent

[14] I am speaking here of *timeless* chances, from which chances at a given time are derived by conditionalising on history up to that time: see Loewer 2001.

possibilities; and you should think that there is a fact of the matter about which of these hypotheses, if either, is true.

Suppose you have just heard the gong ring. In that case, Cr($H|F$) will depend on how you divide our credence between these hypotheses about the chances: it is 1 if you are certain that the chances are “coin first”, 1/2 if you are certain they are “gong first”, and something in between if you assign positive credence to both hypotheses. So unless your credence in “gong first” is 0, the argument from evolving credences must fail. The problem is with CONSTANCY, which will be false, since there is a relevant difference between your centred evidence today (*the gong has just rung*, or maybe *I seem to have heard the gong ringing a moment ago*) and the centred evidence you will have tomorrow (*the gong rang yesterday* or *I seem to remember the gong ringing yesterday*).

This qualification extends to any version of our case where your evidence includes some qualitative property concerning which you have nonzero credence that it distinguishes today from every other day. As with the ringing of the gong, the true objective chance function could in principle treat any such property in either of two different ways. Cr($H|F$) and Cr($H|P$) will depend on how you divide your credences between these hypotheses about the objective chances. To eliminate this potential complication, we can elaborate the setup so as to entail that you have no such distinguishing evidence about today. For example, we can stipulate that your part of the universe (including everything you can have evidence about, but not the Coin) is subject to two-way eternal recurrence.

If we make this further stipulation, CONSTANCY becomes easier to support. Indeed, we might as well stipulate that the recurrence time is one day, in which case your centred credences will clearly have to be the same each day. In that case UPDATING will end up carrying more of the burden of the argument. Someone might argue that UPDATING should fail in the eternal-recurrence scenario, on the grounds that such recurrence would have to involve your losing evidence, and the kind of deference to our future selves that motivates UPDATING does not apply when we think we will lose evidence. It is true the eternal-recurrence scenario is much easier to imagine if we add some process whereby you forget many of the details of your life from one day to the next. The problem with this as a defence of UPDATING is that the details that you are forgetting are not in any way relevant to questions about how the Coin lands—*ex hypothesi* you have no evidence that bears on these questions. It is easy to see how knowing that you will lose some *relevant* evidence could make it rational for you to expect your future beliefs to be inaccurate; it is much harder to see how knowing that you will lose some *irrelevant* evidence could be of any help.

7 Whither the Principal Principle?

If we accept ONE, we must somehow fine-tune our understanding of the connection between chance and credence to explain how the obvious argument for HALF goes wrong. How might this work? Here is a tentative thought. The kinds of things that have chances are coarsely-individuated propositions (e.g. Russellian propositions, or sets of worlds). By contrast, to the extent that credence can be treated as a relation to propositions at all, it is in the first instance a relation to finely-individuated propositions, or modes of presentation of coarsely-individuated propositions. And this disconnect already requires some fine-tuning in the statement of the Principal Principle. It is not true for every mode of presentation of a coarsely-individuated proposition that you must, under that mode of presentation, accord it a conditional credence of x given that its chance is x. Hawthorne and Lasonen-Aarnio (2009) give an example that we can take to illustrate this point. They imagine introducing a name, 'Lucky', by stipulating that it refers to whoever is going to win a certain fair lottery. We can rationally be very confident that Lucky will win the lottery, even though we are also confident that the current chance that Lucky will win the lottery is low. The mode of presentation we associate with the name 'Lucky' is not the right kind of mode of presentation for applying the Principal Principle. Perhaps the lesson of our puzzle is that the 'today' mode of presentation is also of the wrong kind. Even though you know, concerning today, that the chance of its being a Heads day conditional on all its predecessors or successors being Heads days is 1/2, you can still rationally have credence 1 that *today* is a Heads day conditional on all the predecessors (or successors) of *today* being Heads days.

But this escape route raises new questions. If even the 'today' mode of presentation is not of the right kind for the Principal Principle to apply, which modes of presentation *are* of the right kind? Almost all of the modes of presentation under which we have attitudes to *de re* propositions about particular objects and times seem to involve something akin to self-locating thought; there is a risk that the considerations that apply to 'today' will extend very far indeed. But what is the point of attaching chances to propositions about particular objects or times at all, if there are no modes of presentation of these propositions under which our credences in them are constrained by our credences about their chances? Perhaps we should take the basic notion of chance—the notion to which the Principal Principle applies—to attach only to purely qualitative propositions, which describe the overall pattern of qualitative properties and relations without purporting to identify any of the particular objects, events or instants of time that instantiate the pattern. In some ways this is an attractive strategy. Nothing I have said generates any obvious worry about the Principal Principle as applied to purely qualitative propositions, entertained under appropriately impersonal modes of presentation. And so long as one satisfies the demands of the Principal Principle for these propositions, and meets whatever rational constraints the resulting attitudes towards qualitative uncentred propositions might impose on one's attitudes towards qualitative centred propositions (e.g. any true indifference principles for self-locating belief), it is obscure what scope there could be for a failure to satisfy additional demands arising from the application of the Principal Principle to *de re* propositions. But the challenges facing this approach are formidable. In the sciences and in ordinary life, claims about objective chance are almost always about the chances of things happening to particular individuals within particular intervals of time. If the basic notion of chance applies only to qualitative propositions, we must somehow show how to reconstruct in terms of it a notion of chance that can apply to *de re* propositions, and explain why something like the Principal Principle can, at least in ordinary cases, serve as a guide to our thought about these *de re* chances. The difficulty of these tasks should not be underestimated.[15]

[15]Thanks to Frank Arntzenius, David Chalmers, Adam Elga, John Hawthorne, Derek Hiller, and Timothy Williamson.

References

Arntzenius, Frank, Adam Elga and John Hawthorne (2004). 'Bayesianism, Infinite Decisions, and Binding.' *Mind* 113: 251–283.

Briggs, Rachael (2009). 'Distorted Reflection.' *Philosophical Review* 118: 59–85.

Dorr, Cian (2002). 'Sleeping Beauty: In Defence of Elga.' *Analysis* 62: 292–296.

Elga, Adam (2000). 'Self-locating Belief and the Sleeping Beauty Problem.' *Analysis* 60: 143–147.

— (2004). 'Defeating Dr. Evil with Self-Locating Belief.' *Philosophy and Phenomenological Research* 69: 383–396.

— (2007). 'Reflection and Disagreement.' *Noûs* 41: 478–502.

— (MS). 'Subjective Probabilities Should Be Sharp.'

Hawthorne, John (ed.) (2005). *Philosophical Perspectives 19: Epistemology*. Oxford: Blackwell.

Hawthorne, John and Tamar Szabó Gendler (eds.) (2010). *Oxford Studies in Epistemology*, volume 3. Oxford: Oxford University Press.

Hawthorne, John and Maria Lasonen-Aarnio (2009). 'Knowledge and Objective Chance.' In Patrick Greenough and Duncan Pritchard (eds.), *Williamson on Knowledge*. Oxford: Oxford University Press.

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

Joyce, James (2005). 'How Probabilities Reflect Evidence.' In Hawthorne 2005, 153–178.

Kadane, Joseph B., Mark J. Schervish and Teddy Seidenfeld (1986). 'Statistical Implications of Finitely Additive Probability.' In Arnold Zellner and Joseph B. Kadane (eds.), *Studies in Bayesian Econometrics and Statistics*, 59–76. Amsterdam: Elsevier.

Levi, Isaac (1974). ‘On Indeterminate Probabilities.’ *Journal of Philosophy* 71: 391–418.

Lewis, David (1979). ‘Attitudes *De Dicto* and *De Se*.’ *Philosophical Review* 88: 513–543.

— (1980). ‘A Subjectivist’s Guide to Objective Chance.’ In R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, volume 2, 263–93. Berkeley: University of California Press.

— (2001). ‘Sleeping Beauty: Reply to Elga.’ *Analysis* 61: 171–176.

Loewer, Barry (2001). ‘Determinism and Chance.’ *Studies in the History and Philosophy of Modern Physics* 32: 609–620.

McGee, Vann (1994). ‘Learning the Impossible.’ In Ellery Eels and Brian Skyrms (eds.), *Probability and Conditionals: Belief Revision and Rational Decision*, 179–199. Cambridge: Cambridge University Press.

Sturgeon, Scott (2010). ‘Confidence and Coarse-Grained Attitudes.’ In Hawthorne and Gendler 2010, 126–149.

van Fraassen, Bas (1984). ‘Belief and the Will.’ *Journal of Philosophy* 81: 235–56.

White, Roger (2005). ‘Epistemic Permissiveness.’ In Hawthorne 2005.

— (2010). ‘Evidential Symmetry and Mushy Credence.’ In Hawthorne and Gendler 2010, 161–186.

Williamson, Timothy (2007). ‘How Probable is an Infinite Sequence of Heads.’ *Analysis* 67: 173–180.