Philosophical Studies (2007) 135:179–197
DOI 10.1007/s11098-007-9073-4

JONATHAN WEISBERG

CONDITIONALIZATION, REFLECTION, AND SELF-KNOWLEDGE*

ABSTRACT. Van Fraassen famously endorses the Principle of Reflection as a constraint on rational credence, and argues that Reflection is entailed by the more traditional principle of Conditionalization. He draws two morals from this alleged entailment. First, that Reflection can be regarded as an alternative to Conditionalization – a more lenient standard of rationality. And second, that commitment to Conditionalization can be turned into support for Reflection. Van Fraassen also argues that Reflection implies Conditionalization, thus offering a new justification for Conditionalization. I argue that neither principle entails the other, and thus neither can be used to motivate the other in the way van Fraassen says. There are ways to connect Conditionalization to Reflection, but these connections depend on poor assumptions about our introspective access, and are not tight enough to draw the sorts of conclusions van Fraassen wants. Upon close examination, the two principles seem to be getting at two quite independent epistemic norms.

1. INTRODUCTION

Assuming that probabilities are the right way to represent rational belief, are there any constraints beyond the probability axioms that a rational agent should satisfy? Most probabilists think that the rule of Conditionalization is one such requirement. But Van Fraassen (1984) famously opts for an extra synchronic constraint instead: the principle of Reflection. Van Fraassen (1995) claims that Reflection is entailed by the more traditional principle of Conditionalization and he draws two morals from this result. First, that Reflection can be regarded as an alternative to Conditionalization – a more

*My thanks to Frank Arntzenius, Barry Loewer, Chris Meacham, John Hawthorne, and an anonymous referee for their helpful discussion and criticism. Special thanks to Bliss Kern for bringing the topic to my attention.

lenient standard of rationality. And second, that commitment to Conditionalization as a standard of rationality can be turned into support for Reflection. Van Fraassen (1999) also argues that, in many cases of interest, Reflection implies Conditionalization, thus offering a new justification for Conditionalization.

My goal in this paper is to clarify the relationship between Conditionalization and Reflection. I argue that neither principle entails the other, and thus neither can be used to motivate the other in the way van Fraassen says. There are ways to connect Conditionalization to Reflection, but these connections depend on poor assumptions about our introspective access, and are not tight enough to draw the sorts of conclusions van Fraassen wants. Taking one principle as a requirement of rationality does not show the other to be one too. Also, since Conditionalization does not entail Reflection, there is no sense in which Reflection is a liberalization of Conditionalization. Upon close examination, the two principles seem to be getting at two quite independent epistemic norms.

2. FROM CONDITIONALIZATION TO REFLECTION

Reflection says that your current opinion should be constrained by those opinions you think you may come to have in the future. This idea can be precisified in several ways. Here's van Fraassen's original (1984) proposal, which he now calls Special Reflection. Let p be your current credence function and p_t your credence function at some future time t. Then, if you are rational,

Special Reflection (SR) For any H and t, $p(H|p_t(H)=x)=x$.

As is often noted, it immediately follows from SR that your current credence in any proposition H is the expected value of your future credence in

$$H : p(H) = \sum_x xp\ (p_t(H) = x).$$

So SR requires that your current credence be constrained by your foreseeable future credences in a very specific way.

Van Fraassen (1995) later suggested something a bit different: that your current opinion about an event need only be spanned by your foreseeable future opinions.[1] Since van Fraassen includes both credences and expected values under the rubric of 'opinion', we can separate this requirement into two principles of General Reflection:

General Credence Reflection (GCR) For any H and future t, $p(H)$ must lie in the span of your forseeable values for $p_t(H)$.

General Reflection (GR) For any random variable X and future t, the expected value of X relative to p must lie in the span of the forseeable expected values of X relative to p_t.

Of course, GR trivially entails GCR since credences can be regarded as expected values ($p(H)$ is always the expected value of H's indicator function), so GR merits the unqualified moniker, *General Reflection*. But for the purposes of discussion, it will be useful to state GCR separately.

So, which principle follows from Conditionalization? According to van Fraassen, Conditionalization directly entails GR which (along with the assumption of 'Luminosity' to be explained below) entails SR. It is the first alleged entailment that I want to contest. In what sense is Conditionalization supposed to entail GR?

> Some (though not I myself) take as a paradigm of rationality the ideal Bayesian agent, who has opinion in the form of precise numerical probabilities, and changes it solely by Conditionalization on evidence. *Such an agent automatically satisfies the General Reflection Principle*. (Van Fraassen, 1995; p. 17)

Now this ought to strike us as an odd thing to say. How could your status as a conditionalizer, a fact about how you will change your beliefs in the future, constrain what you believe today? Here is what van Fraassen has to say about it:

> Starting with probability function p now he [the conditionalizer] will have at time t one of the functions $p(\cdot|E(i,t))$ where $E(i,t)$ is a possible evidence scenario between now and t. Because $E(i,t)$ is a partition (disjoint and exhaustive), probability theory entails that $p(H)$ is a convex combination

of, hence lies in the interval spanned by, the numbers $p(H|E(i,t))$. (Van Fraassen, 1995; p. 17)

Now you might well ask why an agent's possible evidence scenarios should form a partition. Let's set that aside for now. Even assuming that they do form a partition, there's a more serious problem here. All van Fraassen has proven is that the agent's current opinion is spanned by those opinions he in fact may have. If he is a conditionalizer, then the functions $p(\cdot|E(i,t))$ are the opinions he may come to have. But what he *thinks* is possible is another matter. This argument says nothing about whether his current opinion is spanned by the opinions he *thinks* he may have, and it's those opinions that matter to Reflection.

So van Fraassen has not shown that an agent who satisfies Conditionalization automatically satisfies GR. Could he be right anyway? First disambiguate two senses in which a Conditionalizer might be thought to satisfy Reflection 'automatically'. First, we might suspect that anyone who will be a strict Conditionalizer from now on satisfies Reflection now. This seems to be what van Fraassen has in mind – he uses your status as a future conditionalizer to show that you satisfy Reflection now. Second, we might hope that anyone who Conditionalizes always satisfies Reflection as a result. That is, we might try to show that the deliverance of Conditionalization is always a Reflective credence function. Not surprisingly though, neither of these 'automatic' relationships holds. It's easy to construct sequences of probability functions, each member obtainable from its predecessor by Conditionalization, such that no member of the sequence satisfies SR. The same can be done for GR. So it seems that van Fraassen is simply wrong that an agent automatically satisfies GR in virtue of obeying Conditionalization.

Still, it would be too strong to say that there is no connection at all between Conditionalization and Reflection being demonstrated here. True, an agent may obey Conditionalization while snubbing Reflection, but in order to do so she must think herself capable of violating Conditionalization.

What van Fraassen has shown is that an agent who is absolutely certain she will always obey Conditionalization automatically satisfies GR. For suppose an agent is absolutely certain she will always conditionalize, in the sense that it is not epistemically possible for her that she will do otherwise. Then her epistemically possible future credences in H are just the $p(H|E(i,t))$. Assume also that she knows her own conditional credences (a non-trivial assumption). Then she also knows the values of the $p(H|E(i,t))$, and so her current credence in H is spanned by the values she thinks she may come to have, for just the reason van Fraassen gives. The moral is that whether an agent satisfies Conditionalization has nothing to do with whether she satisfies Reflection. What matters is whether she is certain she will obey Conditionalization. What van Fraassen has shown is: absolute certainty that one is a conditionalizer implies GR-satisfaction, assuming you know your own conditional credences.

Well, almost. Two concerns need to be addressed before we can accept this result, one major and one minor. The minor concern: van Fraassen showed that, if an agent thinks she is a Conditionalizer, then her current credence is spanned by the credences she thinks she may come to have. But this only gives us GCR, whereas we want the full principle of GR; not just for its own sake, but also because van Fraassen uses the full GR to derive SR.[2] Fortunately, it's not difficult to turn van Fraassen's proof of GCR into a proof of GR,[3] so we can set this concern aside.

More serious is the worry I bracketed just a moment ago. Van Fraassen's proof assumes that the 'evidence scenarios' our ideal Bayesian may encounter (the $E(i,t)$) will form a partition. But why should that be? Presumably, between today and tomorrow I could learn any number of individual facts $E_1, \ldots, E_n$, as well as many combinations of those facts. In that case many of my possible evidential scenarios are not exclusive, $E_1 \wedge E_2$ and E_1 for example. So why does van Fraassen assume that they are? Worse yet, the scenarios might not be exhaustive either. Suppose I think that, in the next day, I

could learn that A won the election or that B did, but not that it was a tie—if there's a tie, I won't learn about it for a while. Then my possible evidential scenarios don't form a partition. I can learn A or I can learn B, but these propositions don't exhaust the space of possibilities. So again: why does van Fraassen assume that evidential scenarios always form a partition?

Here's my guess: the partitioning assumption results from a confusion between the epistemic paths an agent may take and the information she learns on those paths. While the possible histories I may encounter between now and t do form a partition of the space of possibilities, the information that I may glean along those histories needn't form a partition. To make this point vivid, we can visualize an agent's epistemic history as a ticker-tape, where each cell of the tape corresponds to a time and contains the information the agent conditionalizes on at that time (the cell is blank if she doesn't learn anything then). Now, the set of possible tapes certainly forms a partition, since an agent must undergo exactly one tape. But the contents of the tapes—the conjunctions of cell-contents—needn't obviously form a partition, for the reasons already given. Confusingly, both a tape and its contents are aptly called an 'evidential scenario', and so it's easy to mix the two up. But it's the tapes that form a partition, while it's their contents that an agent conditionalizes on. So if the $E(i,t)$ are what the agent may conditionalize on, they needn't form partition.

Unless, that is, we can find some way to equate tapes with their contents. To defend his argument, van Fraassen might respond by insisting that what an agent conditionalizes on is not just the contents of a tape, but also the fact that she encounters that tape. Then his argument would be free of equivocation. As it happens, an assumption that van Fraassen later uses to derive SR from GR yields just this result. Call it,

Luminosity For any H,t, if $p_t(H) = x$ then $p_t(p_t(H) = x) = 1$.[4]

Intuitively speaking, Luminosity says that an agent always knows her own credences. This implies that the $E(i,t)$ form a partition as follows. Suppose our agent is always Luminous and is a Conditionalizer. Then whenever she learns some fact

E at t, she knows thereafter that $p_t(E) = 1$ at that time and not before. So she knows her own evidential history, i.e. she knows which ticker-tape she has been reading. Since the tapes are exclusive and exhaustive, so are the contents one learns when reading them. Thus Luminosity implies that a conditionalizer's $E(i,t)$ form a partition.

Admittedly, this connection is a bit surprising. Why should an assumption about introspective access yield a result about the sort of evidence you can get? The answer lies partly in an assumption implicit in our notation, and partly in the perfect recall required by Conditionalization. As formulated, Luminosity implicitly assumes that the agent always knows her current credence under a *de dicto* temporal description—she doesn't know that she has credence x in H 'now', but that she has it 'at time t'. Since Conditionalization ensures that she never forgets these *de dicto* facts, she assembles a perfect record of what she learned when as she goes. This might not happen, of course, if she were luminous in a *de se* way, for then it's not clear how Conditionalization applies. If an agent gets no new evidence between now and t, Conditionalization says that she should leave her credences unchanged. But how should this apply to an irreducibly indexical hypothesis like "P holds now"? Since the dynamics of belief for such hypotheses is an open and tricky question, we have to make do with purely *de dicto* resources. An artifact of this limitation is that our best formulation of Luminosity leads to partitioning for conditionalizers.

It's worth noting that we've seen something of Luminosity already. I said van Fraassen's proof shows that if you're certain you will conditionalize and you know your conditional credences, you satisfy GR. The assumption that you know your conditional credences can be gotten by assuming that the agent is luminous. However, assuming that the agent is luminous isn't quite what we need to get partitioning and close the last hole in van Fraassen's proof. The logically possible $E(i,t)$ of a luminous conditionalizer do form a partition but we need an agent whose *epistemically* possible $E(i,t)$ form

a partition. So we don't need to assume that the agent will be luminous in the future. Instead, we want to assume that she is certain that she will be luminous, so that her epistemically possible $E(i,t)$ form a partition. To summarize our results so far then, van Fraassen's alleged proof that one satisfies GR in virtue of being a conditionalizer fails. But we can show that if (i) you are certain you will always conditionalize, (ii) you are luminous now (at least with respect to your conditional credences), and (iii) you are certain you will be luminous in the future, then you satisfy GR.[5]

3. EVALUATION OF THE REVISED RESULT

Could this revised result still support the morals van Fraassen sought to draw? It seems clear enough that Reflection is in no way a generalization of Conditionalization. Conditionalization is a diachronic constraint; it says what sequences of probability distributions are permitted. A more liberal policy would do the same thing but allow a superset of those sequences. While Reflection can also be seen as a constraint on what probability sequences are allowed (those that obey Reflection throughout), the allowed set crosscuts the one allowed by Conditionalization. The distributions allowed by Reflection are neither a superset nor a subset of those allowed by Conditionalization. Still, van Fraassen has shown that the set of distributions that meet (i)–(iii) above are a subset of the Reflective ones. So Reflection can be seen as a liberalization of a principle that takes (i)–(iii) to be requirements of rationality. The trouble is that no such normative principle can be correct. Certainly we oughtn't think that we will always conditionalize since we have overwhelming evidence that we rarely do. So (i) is not a normative requirement. The same goes for (iii). We clearly aren't luminous, so surely we ought not be certain we are. As for (ii), while Luminosity does describe an ideal that it would be nice to live up to, I'll argue in section 4 that it can't be regarded as a norm of rationality.

What about the other conclusion van Fraassen wanted to draw from his argument, that Conditionalization can be used to support Reflection? Presumably the intended reasoning was something like: if you ought to obey Conditionalization, and by obeying Conditionalization you automatically satisfy Reflection, then you ought to obey Reflection too. Any violation of Reflection is a violation of Conditionalization and hence an irrationality. But we've seen that that's not really so. You can go your whole life without obeying Reflection and still not violate Conditionalization. And we can't use parallel reasoning with (i)–(iii) in place of Conditionalization, since they don't describe norms of rationality. If you violate Reflection you will violate at least one of (i)–(iii), but that doesn't mean you're irrational. So this sort of reasoning doesn't offer any support for Reflection as a principle of rationality.

A third motivation van Fraassen seems to have in connecting Conditionalization to Reflection is to redirect all attacks on Reflection towards proponents of Conditionalization. In his own words, "It is wonderfully remarkable and disturbing that all the criticisms directed at the Reflection Principle were not already previously raised. What was more salient in the literature than the Bayesian principle that the ideal epistemic subject updates his opinion by Conditionalization? As we have just seen, the one implies the other." (1995, p. 17) Indeed, van Fraassen seems to be on to something here. The infamous case of Sleeping Beauty is often thought to be one where Conditionalization and Reflection are violated together. If Beauty begins with credence 1/2 in heads and ends up with credence 2/3, she violates Reflection since she knows ahead of time that she will have credence 2/3 in heads. She also seems to violate Conditionalization, since she receives no evidence in the interim. Also, Arntzenius (2003) offers five cases where it seems we ought to violate both Reflection and Conditionalization. Admittedly it's a bit suspicious that the two principles ought so frequently to stand or fall together if they really are independent as I'm claiming. So an explanation is in order.

Why are violations of Reflection so often violations of Conditionalization? Because the cases given are typically ones where the agent not only violates Conditionalization, but knows that she will. In the cases described by Elga (2000) and Arntzenius (2003), the violations of Reflection are obtained by considering an agent who foresees two possible futures, both of which lead her to the same credence (different from the one she has now). Hence she violates Reflection. Assuming Luminosity, we then know that she cannot believe that she is a Conditionalizer. Assuming also that her beliefs are correct in this respect, we get a case where she violates Conditionalization too. The cases in question violate Conditionalization because we take the agent's beliefs about how she will proceed in her possible futures to be correct. In fact, we needn't do this in order to obtain violations of Reflection in cases like Arntzenius' and Elga's. Sleeping Beauty doesn't actually need to undergo the experiment in order for her to reasonably violate Reflection. All that is required is that she reasonably believe she will.

To illustrate this analysis, consider Arntzenius' Shangri-La case. Suppose the deities have granted you a visit to Shangri-la, but they require that no one who comes to Shangri-la know how they got there. So they decide to take you on one of two paths, the one by the mountains or the one by the sea. The choice is to be decided by a fair coin-flip: the sea if heads and the mountains if tails. If you do go by the mountains, however, a spell will be cast when you enter the gates of Shangri-la and you will remember having gone by the sea. So either way, you will remember having gone by the sea. Now suppose that, as it happens, the coin comes up heads and so you really do go by the sea. En route, you are certain that the coin came up heads. But when you arrive at the gates of Shangri-la, you drop your credence in heads to 1/2 since, for all you know, your memory of having traveled by the sea is fictitious. Your degrees of belief in this case violate Reflection since, while traveling by the sea, you are certain that your future credence in heads will be lower than it actually is.

Nevertheless, you are rational. Interestingly, you violate Conditionalization too. When you arrive at the gates of Shangri-la you gain no new evidence since nothing happens that you did not foresee. Nevertheless, you change your credence in heads from 1 to 1/2.

Thus we have a case where a violation of Reflection is accompanied by a violation of Conditionalization. But notice that the violation of Reflection happens en route while the violation of Conditionalization happens upon your arrival. We could have stopped the story at the sea and gotten our violation of Reflection without worrying about what happens next. Indeed, if you had gone on to stick to your guns upon arrival we would have had a violation of Reflection without violating Conditionalization. As with Sleeping Beauty, the two needn't go together. The reason the violation of Conditionalization does happen in Arntzenius' case is that, in the natural telling of the story, you actually do what you think you will do: namely undergo a non-conditionalizing shift. Since you violated Reflection you had to think you would do that all along, and the natural way to tell the story is that you were right about that. But that's not the only way to tell it.

4. OTHER CONNECTIONS?

We've seen that Conditionalization isn't enough to ensure satisfaction of Reflection. Might there still be some way to draw a connection? We could try a couple of things here: we could try to show that Reflection follows from Conditionalization with the help of added assumptions, or we could try to show that Conditionalization entails some Reflection-like principle that suffices to capture the intuitions that motivate Reflection. I'll consider two such attempts now.

4.1. *Adding an Assumption*

What assumption could we add to Conditionalization to recover Reflection? A superficial resemblance between

Reflection and Lewis' Principal Principle (Lewis, 1980) is suggestive in this connection. Finessing certain complications, that principle is:

Principal Principle (PP) For any H and t, $p(H|c_t(H) = x) = x$, where $c_t(H)$ is the objective chance of H at t. PP is partly motivated by the truism that an agent who believes at t that the chance of H at t is x, ought to be sure to degree x that H. Assuming that rational learning is just Conditionalization, PP then seems a natural constraint. If you were to violate it, you might learn that $c_t(H) = x$ and come to have some credence in H other than x. Presumably someone in that position violates some sort of conceptual coherence. Part of what it is to believe the chance of such-and-such is x is to think that you ought to set your credence that such-and-such accordingly. So if we assume Conditionalization to be the sole method for rational updating, PP looks like an appropriate formalization of one intuitive connection between chance and credence.

Now, since PP is basically just SR with c_t in the place of p_t, it's natural to ask whether Conditionalization provides a similar motivation for SR. If you violate SR, i.e. you have $p(H|p_t(H) = x) \neq x$ for some H and future t, then conditionalizing on $p_t(H) = x$ at t will yield $p_t(H) \neq x$. This violates a requirement we might call.

Transparency If $p_t(p_t(H) = x) = 1$ then $p_t(H) = x$.

Transparency is Luminosity's converse, and says that you are never wrong about your credences when you are sure of them. In fact, Luminosity implies Transparency[6] but not vice versa. So this way of connecting Conditionalization to Reflection might be seen as an improvement over van Fraassen's argument, since it employs the strictly weaker assumption of Transparency. Why respect SR? Because otherwise Conditionalization may lead you into a violation of Transparency.

We might spurn this argument for its appeal to a very strange sort of evidence. The argument considers a possible future in which you gain as evidence a fact about what your credences are about to be. At t, you learn that $p_t(H) = x$.

This evidence has a weirdly self-fulfilling character (or self-defeating, depending on your priors), and one might object that such evidence is not possible. I confess, though, that I don't find this move very compelling. After all, what's to stop an oracle from telling you what your credence is about to be?

I think a more moving criticism is that Transparency is a poor assumption. While Transparency may describe an ideal that it would be nice to live up to — it would be nice to be right about one's own credences just as it's nice to be right about anything—it doesn't describe a norm of rationality. The chance-credence truism behind PP may be supported by some definitional feature of the chance concept, but someone who wrongly thinks they have credence x isn't suffering from any conceptual incoherence. They don't fail to grasp what it is to have credence x, they are just wrong about their own psychology. Transparency requires infallibility in a contingent, empirical domain, and failure to live up to such a requirement, while unfortunate, does not make for irrationality.

Well, fair enough, we shouldn't say that an agent is always irrational in virtue of violating Transparency. But isn't she irrational if she violates Transparency when she could have avoided it? SR is a constraint on priors that prevents just this sort of eventuality: if you satisfy SR and you are a strict Conditionalizer, you avoid violating Transparency in cases where your evidence is $p_t(H) = x$. This doesn't require any special introspective insight or empirical infallibility, it just requires that you organize your priors in a particular way. Given that you can use SR as an a priori safeguard against certain violations of Transparency, why shouldn't you?

Because using SR as a guard against violations of Transparency comes with a price. There are lots of a priori safeguards against error, most of which we do not think are good policy. To guard against believing explicit contradictions I might simply never believe a conjunction, but it certainly doesn't follow that this policy is a norm of rationality. In general, the directive to adopt policies that prevent error must be conditional. One should only adopt a policy as an a

priori safeguard against error when the benefits of avoiding that error outweigh the costs of adopting the policy. And SR does have costs. A reflective agent treats all evidence as trumpable by evidence about her future credences. Thus she pays the price of allowing her future credences to dictate her current credences regardless of what evidence she receives in the meantime. In doing so, she makes herself vulnerable to scenarios where she adopts a credence for the sole reason that she thinks she will, regardless of the other evidence at hand. If I learn that in a moment I'll see purple elephants, Reflection will require that I adopt that credence now, even though it seems I shouldn't. Adopting that credence right now will ensure that I'm right about what I'll think in a moment, but it will also ensure that I'm disastrously wrong about the purple elephants, since there aren't any. Using Reflection to insure that your second-order beliefs are correct works by bringing your first-order beliefs in line with the second-order ones. The price you pay with this method is that your first-order credences are at the mercy of your beliefs about them —even to the exclusion of intuitively good evidence to the contrary. It may be good policy to arrange for your first-order credences to be brought in line with your second-order ones when your second-order beliefs foresee first-order beliefs that will be formed for good reasons. But it can't be a good policy universally.

This problem illustrates a general problem with arguing from ideals to norms. The Transparency-based argument does show that ideal agents obey Reflection but it fails to show that we ought to obey Reflection. In general, showing that X holds in ideal scenarios does not imply that we ought to aspire to X. What ideals we ought to aspire to depends on what limitations we face. Even if ideal agents always do X, it may not be a good idea for us to do it because our situation is less than ideal. Of the possible outcomes that are attainable for us, X-outcomes may be less than ideal.

4.2. *Capturing the Intuition Behind Reflection*

A common objection to SR is that it's too strong. SR requires you to adopt a credence today if you know you will have it tomorrow, regardless of why you will have it tomorrow. Sometimes that means knowingly pursuing tomorrow's irrationality today. Nevertheless, SR does have at least some intuitive grounding: if you know you're going to have credence x tomorrow and you know that this state will come about rationally, why not have it today? Presumably your later credences are supposed to be improvements over your earlier ones (otherwise why change them?), so why wait to make the improvement when you know what it'll be? SR may be too strong but it seems we do want some such principle; something like, "if you believe you'll have some credence tomorrow rationally, you should have it today." But even that may be too strong. Loss of information due to memory loss or loss of self-locating information seems to be rationally permissible, yet we shouldn't preemptively adopt the resulting credences before the loss of information happens. Our reflective intuitions only apply to cases where the foreseeable credence isn't just rational, but is in some sense a strict improvement over our current state. One way this might happen is if our future state is a strict gain in information, and this is where Conditionalization seems to be helpful.

Suppose I find out that tomorrow I'll have credence x in H as the result of learning some fact E. I may not know which fact I'll learn, but I do know that I'll learn one of the Es for which $p(H|E) = x$. If we let $E_1 \vee \cdots \vee E_n$ be the disjunction of all such propositions, then I should conditionalize on that disjunction. Assuming that the E_i are mutually exclusive, this will give me credence x now since, as a matter of the probability calculus, $p(H|E_1 \vee \cdots \vee E_n) = x$ when $p(H|E_i) = x$ for all i. Thus Conditionalization seems to account for a fundamental intuition behind Reflection. Why think that we ought to sometimes adopt our foreseeable credences? Because the credence will be based on one of a collection of pieces of

evidence and, even if you don't know which piece of evidence you'll get, you've in effect learned their disjunction. And that's enough to ground that same credence via Conditionalization.

This story is questionable on a few points—why the E_i should be exclusive, for example—but the really troublesome bit is that it implicitly assumes Luminosity. The explanation may work for an agent who knows which E_i she may learn tomorrow, but what about an agent who doesn't know which Es are such that $p(H|E) = x$? She knows that she'll learn some such proposition, but she can't infer from that existentially quantified, meta-linguistic fact to the ordinary disjunction $E_1 \vee \cdots \vee E_n$ unless she knows her own conditional credences for H. That is, unless she is at least partly luminous. Our reflective intuitions aren't restricted to such agents, however. Even if I have no idea which propositions I treat as warranting a credence of x, I still think I ought to adopt credence x now given that I'm going to learn one of them tomorrow. The bottom line is that learning $p_t(H) = x$ doesn't motivate $p(H) = x$ by telling us some more mundane information, like the disjunction of the E_i. Rather, we seem to think that a credence of x is required today based solely on the fact that we will have it tomorrow as the result of learning *something*. SR may be overkill as an attempt to formalize such reflective intuitions, but Conditionalization won't do the job either. Whatever the appropriate formulation of our reflective intuitions looks like, it seems it will have to go beyond Conditionalization.

5. FROM REFLECTION TO CONDITIONALIZATION

So much for deriving Reflection form Conditionalization. What about the other way round? Van Fraassen (1999) argues that General Reflection offers a new justification for Conditionalization since, at least in many cases of interest, GR implies Conditionalization. That argument makes a

mistake analogous to the one pointed out in Section 2 and, as a result, only ends up showing that Reflective agents will *think* they will Conditionalize, though they may not. I'll close by reviewing this argument and pointing out the faulty step.

Roughly speaking, GR is supposed to require Conditionalization when the agent is certain that her evidence at t will be one of the elements of a partition. To make this condition precise, let $\{E_i\}$ be the partition in question and let q_i be the distribution the agent thinks she will come to have when she receives E_i as evidence at t. The precise statement of the condition is then

Condition $\{q_i\}$ is a set of probability functions such that $q_i(E_i) = 1$ for each E_i.

Condition captures the assumption that exactly one of the E_i will be treated as evidence, i.e. will be given credence 1. Letting p represent the agent's current distribution, and assuming Condition, van Fraassen shows

Result If for every random variable X, $E(X,p) \in \text{Span}[X,q_i]$, then $p(\cdot|E_i) = q_i$ for every i such that $p(E_i) > 0$.

Since a reflective agent satisfies the antecedent of Result by definition, this is supposed to show that she will conditionalize when the evidence comes in. Whichever E_i she receives, she will adopt q_i as her new distribution, and q_i just is $p(\cdot|E_i)$ by Result.

Given the discussion in Section 2, the problem here should be fairly apparent. While it's true that a reflective agent will satisfy Result's antecedent, it doesn't follow that she will conditionalize. It follows that her q_i are the same as the $p(\cdot|E_i)$, but the q_i are just the distributions she thinks she will come to have when the evidence E_i comes in. What she will actually do is another story. This flaw in the argument shouldn't be surprising since van Fraassen is trying to show that you will conditionalize merely by looking at your current degrees of belief. It would be very strange if your current epistemic state placed logical limitations on your future state.

Could we show something a bit weaker? Perhaps that an agent who is not only reflective now but throughout her life

always conditionalizes? That would entail that any two reflective distributions are always relatable by Conditionalization, which is easily shown to be false by counterexample. As with van Fraassen's attempt to derive GR from Conditionalization, we must settle for a modified result: a reflective agent who satisfies Condition should think that she will conditionalize, though she needn't actually do so.

NOTES

[1] A value is spanned by a set of values iff it can be obtained as a mixture of the set's elements.

[2] See van Fraassen (1995), pp. 18–19, for the proof that GR entails SR.

[3] Van Fraassen's argument shows that, if your foreseeable future credences are the $p(H|E(i,t))$, then not only is $p(H)$ a mixture of the foreseeable $p_t(H)$, but the entire function p is a single mixture of the possible future p_t. Let that mixture be $p = \sum_i x_i p_{t,i}$. Then, for any random variable X, its expected value relative to p is

$$\begin{aligned} E(X,p) &= \sum_j X_j\, p(j) \\ &= \sum_j X_j \sum_i x_i\, p_{t,i}(j) \\ &= \sum_i x_i \sum_j X_j\, p_{t,i}(j) \\ &= \sum_i x_i E_i\ (X,p_t). \end{aligned}$$

Here $E_i(X,p_t)$ is the expected value of X relative to p_t, supposing you undergo evidential scenario i. Thus your current expected value for X is a mixture of your foreseeable expected values for X.

[4] I'm borrowing 'Luminosity' from Williamson (2000), though my use of it differs a bit. Agents who are luminous in Williamson's sense are not only certain of their credences, they *know* them. And they know all their other mental states too.

[5] My thanks to an anonymous referee for pointing out the need to separate (iii) from (ii).

[6] Assume an agent satisfies Luminosity and is coherent. Coherence implies that, if she has $p_t(p_t(e) = x) = 1$, then $p_t(p_t(e) = y) = 0$ for any $y \neq x$. So Luminosity implies that none of these other y values is correct, i.e. $p_t(e) = x$.

REFERENCES

Arntzenius, F. (2003): 'Some Problems for Conditionalization and Reflection', *Journal of Philosophy* C.7.

Elga, A. (2000): 'Self-Locating Belief and the Sleeping Beauty Problem', *Analysis* 60.

Lewis, D. (1980): 'A Subjectivist's Guide to Objective Chance.' *Studies in Inductive Logic and Probability*, II.

Van Fraassen, B. (1984): 'Belief and the Will', *The Journal of Philosophy* 81.

Van Fraassen, B. (1995): 'Belief and the Problem of Ulysses and the Sirens', *Philosophical Studies* 77.

Van Fraassen, B. (1999): 'Conditionalization: a New Argument For', *Topoi* 18.

Williamson, T. (2000): *Knowledge and its Limits*, Oxford University Press.

Rutgers University
26 Nichol Avenue
New Brunswick, NJ 08901,
USA
E-mail: jweisber@rci.rutgers.edu