Conditionalization Without Reflection[*]

Jonathan Weisberg
Rutgers University

Abstract

Conditionalization is an intuitive and popular epistemic principle. By contrast, the Reflection principle is well known to have some very unappealing consequences. But van Fraassen argues that Conditionalization entails Reflection, so that proponents of Conditionalization must accept Reflection and its consequences. 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. I also propose a replacement for Reflection that accounts for the intuitions that made Reflection appealing, but doesn't lead to Reflection's bad consequences.

1 INTRODUCTION

I'm going drinking tonight. By closing time I expect to have drunk enough for my judgment to be impaired. I'll systematically overestimate my own abilities, especially my ability to drive safely. Knowing this about my future self, should I now adopt my future opinion and believe that I will be competent to drive home tonight? Surely not, and any principle that says I should is one we'd have to reject. The Reflection principle, which requires that my current beliefs match the ones I expect to have in the future, is such a principle (Talbott 1991).

Now here's a plausible story about how I ought to change my degrees of belief when I learn something. Suppose I'm playing three card monte and I have no idea which card is the queen; for each card, my confidence that it's the queen is 1/3. Luckily, the house slips up and I catch a glimpse of the middle card, ruling it out. My confidence for each of the two remaining cards should now be 1/2. In general, when I eliminate a possibility I should redistribute my credence as follows: for each remaining possibility, divide its old credence (1/3) by the amount of total credence left over (2/3). That is, obey the principle of Conditionalization.

[*] This paper is largely based on my paper "Conditionalization, Reflection, and Self-Knowledge", forthcoming in *Philosophical Studies*. My thanks to Frank Arntzenius, Adam Elga, John Hawthorne, Barry Loewer, Chris Meacham, Adam Sennet, and an anonymous referee for *Philosophical Studies* for their helpful discussion and criticism.

I am for Conditionalization and against Reflection. Conditionalization is an intuitive and powerful principle, accounting for our judgments about good reasoning in all kinds of cases. But Reflection requires us to adopt our future credences even when we think they will be badly mistaken, as in the drinking case. This would be a happily stable position, except that van Fraassen (1995) argues that Conditionalization actually entails Reflection. If he is right then I, and a lot of other people, are in trouble. If we want to have Conditionalization, we have to take Reflection along with it, nasty consequences and all.

My main aim in this paper is to get us out of this bind. My strategy is to argue that van Fraassen is simply wrong that Conditionalization entails Reflection, so we are free to endorse Conditionalization without Reflection. In section 2 I present my argument that Conditionalization does not entail Reflection, and in section 3 I anticipate some objections. In Section 4 I consider other possible connections between Conditionalization and Reflection. To tighten my case, I rule out an alternative argument that uses Conditionalization to motivate Reflection by analogy with the Principal Principle. Taking a brief detour, I also refute Van Fraassen's (1999) argument that Reflection offers a new justification of Conditionalization. In Section 5, I propose a replacement for Reflection, designed to account for the intuitions that motivated Reflection without inviting the problems. I conclude that Conditionalization without Reflection is not only a logical possibility, it's actually a very attractive position.

Before I get into all that though, I want to say something about where I'm coming from here. You might wonder how anyone could be for Conditionalization and against Reflection, given that dutch book arguments can be given for both. Dutch book arguments show that, if you violate a given principle, you can be suckered into a set of bets that *necessarily* lead to a net loss. Since you can see this disaster coming even before taking the bets, dutch bookability is supposed to signify irrationality. So how can I be for Conditionalization but not Reflection when the same arguments are standardly given for both principles? The answer is that I don't find dutch book arguments convincing. In the last 20 years or so, dutch book arguments have come under a lot of fire. While they used to be regarded as pretty decisive, they are now commonly regarded as question-begging or invalid. There is a lot of critical literature on dutch books, so I won't discuss the issue here; my discussion just presumes that dutch book considerations have been ruled out. For some critical discussions of dutch book arguments, see (Schick 1986), (Earman 1992), and (Weatherson 1999).

How do we argue for or against principles like Conditionalization and

Reflection if we give up the dutch book approach? One thing we can do is evaluate how well a principle gels with paradigmatic cases of good and bad reasoning. Considerations like the three card monte game and the drinking case are supposed to motivate my favored view in this way, since Conditionalization looks good and Reflection looks bad. Of course, it takes a lot more than just a couple of examples to make the case. But years of Bayesian epistemology and philosophy of science show us that something like Conditionalization has got to be right, whereas Reflection seems to have almost no use at all, and even gives disastrous advice in mundane cases like the drinking example. As I said, van Fraassen has tried to get some use out of Reflection by using it to offer a new justification of Conditionalization, but I show in section 4.2 that his argument there is mistaken. And, while there are cases where Reflection looks good, section 5 shows that we can account for those cases by appealing to a weaker, less troublesome principle. Given all this, it would be best if we could have Conditionalization without Reflection. I'll now argue that we can.

2 van fraassen's argument

Reflection says that your current opinion should be constrained by the opinions you think you might come to have in the future. Constrained how? Informally put, van Fraassen's original (1984) proposal was this: your credence in a proposition H, given that your credence in H will be x in the future, should be x right now. To make this precise, let p be the probability function representing your current degrees of belief, and p_t your credence function at time t. Then, if you are rational,

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

The notation $p(A|B)$ is standard shorthand for $p(A \wedge B)/p(B)$, and is usually read 'the probability of A given B'. In the informal statement above I'm using 'given' in this technical sense.

To get a handle on Special Reflection, it helps to apply one of its immediate consequences to an example. From SR and a bit of arithmetic, we can show that your current credence in any proposition H should be the expected value[1] of your future credence in H: $p(H)=\sum_x x\,p(p_t(H)=x)$. So in the drinking case, if you are 2/3 confident right now that you will be 8/10 confident later that you can drive safely, and 1/3 confident that you will be 9/10 confident

[1] The expected value of a random variable X is the sum of the products of its possible values and their probabilities: $p(X_1)X_1 + p(X_2)X_2 + \ldots + p(X_n)X_n$, or $\sum_i p(X_i)X_i$ for short.

later that you can drive safely, your current degree of belief that you can drive safely should be $2/3(8/10) + 1/3(9/10) = 5/6$. Thus SR requires that your current credence be constrained by your foreseeable future credences in a very specific way. From your foreseeable levels of confidence in H, and your current confidence that you will have those levels of confidence, you can derive what your confidence in H should be right now.

Van Fraassen (1995) later suggested something a bit different: that your current opinion about a proposition need only be *spanned* by your foreseeable future opinions. For real numbers, being spanned by a set just amounts to lying between its highest and lowest values. So this requirement is more lenient than SR. SR not only requires that your credence in a proposition lie between the highest and lowest foreseeable future values, but also that it be the expected value of the foreseeable values. This latter requirement is now being dropped. But, something is being added too. Van Fraassen is using 'opinion' liberally here, to mean expected values in general. This includes degrees of belief, since your degree of belief in A is just the expected value of A's indicator variable[2], but it applies to other expected values too. The official statement of the principle is thus:

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.

To take an example, suppose you're wondering how well your favorite soccer team will do in their game this weekend. One thing GR requires of you is that your current degree of belief that they will win be spanned by the degrees of belief you think you may come to have at half-time. If you know that you are always optimistic at half-time — between 1/2 and 9/10 confident, say — then your confidence should be between 1/2 and 9/10 now. But also, GR requires that your expected value for the number of goals your team will score lie between the expected values you think you might come to have at half time. If you know that, at half-time, your expected score for your team will between 3 and 5, then it had better be in that interval now.[3]

[2] A proposition's indicator variable is the variable that is 1 when it is true, 0 when it is false.

[3] You might think this second requirement is automatically satisfied whenever you satisfy the first, but this isn't so. To take a simple example, consider a random variable X that can take values in the vector $\langle 1, 2, 3\rangle$, and suppose your credences in those values are given by $\langle 1/3, 1/3, 1/3\rangle$. Then your expected value for X is 2. If your foreseeable credence distributions are $\langle 0, 2/3, 1/3\rangle$ and $\langle 1/3, 0, 2/3\rangle$, then your foreseeable expected value for X is $7/3$ in either case.

So, which version of Reflection 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 $p(\cdot|E(i,t))$[4] where $E(i,t)$ is a possible evidence scenario between now and t. Because $E(i,t)$ is a partition[5] (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? Let's disambiguate

[4] Regarding notation: $p(\cdot|E(i,t))$ is just the probability function you get by conditionalizing p on $E(i,t)$. That is, the probability function that assigns $p(H|E(i,t))$ for every proposition H.

[5] A partition is a set of propositions that are mutually inconsistent, but whose disjunction is a tautology, $\{A, \neg A\}$ for example. In possible world terms, a partition carves up the space of worlds into distinct pieces, like a cake.

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.[6] So the fact that you're going to conditionalize doesn't mean you satisfy Reflection right now, nor does it mean that you will after you've conditionalized. Van Fraassen is simply wrong that an agent automatically satisfies Reflection 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 GR requires

[6] It's well known that, given any two probability distributions, the space of propositions can be enriched in such a way that the second will now be obtainable from the first by Conditionalization. So all we have to do is pick two distributions that violate Reflection and apply this result. Ironically, a proof of it appears in (van Fraassen 1989: p. 322)

that all random variables, not just credences, be spanned by the foreseeable values. This is especially important given that van Fraassen uses the full GR to derive SR.[7] But it's not difficult to extend van Fraassen's proof into a proof of GR proper,[8] 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' a person may encounter (the $E(i,t)$) will form a partition. But why should that be? 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 parition. 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 a

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

[8] 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 $\sum_j X_j p(j) = \sum_j X_j \sum_i x_i p_{t,i}(j)$. Rearranging terms, we have $\sum_i x_i \sum_j X_j p_{t,i}(j)$ which is the same as $\sum_i x_i E_i(X, p_t)$. 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.

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 and t, if $p_t(H) = x$ then $p_t(p_t(H) = x) = 1$.[9]

Intuitively speaking, Luminosity says that an agent is always aware of 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

[9] I'm borrowing 'Luminosity' from Williamson (2000), though my use of it differs a bit. For Williamson, if your credences are luminous then you are always in a position to know what they are, though you might not actually know. If you do know, they will have evidential probability 1 for you. But that's not the same as certainty; 'evidential probability' is a term of art not to be confused with 'degree of belief'.

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. 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. Recall that Reflection, whether Special or General, says that your current credences should be constrained by the credences you *think* you may come to have. So we need to ensure that the evidential scenarios you *think* you may come to have form a partition. To get this result, we have to assume that you *think* will be luminous, so that your epistemically possible $E(i,t)$ form a partition. Then, if you also think you will conditionalize, we get that your foreseeable credences are the ones van Fraassen says they are, the $p(\cdot|E(i,t))$.

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.

Could this revised result still support the moral van Fraassen wanted to draw? If Conditionalization is a norm of rationality, does this show that Reflection is too? The originally 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. Now, we've seen that that's not really so. You can go your whole life without obeying Reflection and still not violate Conditionalization. But van Fraassen has shown that if you satisfy conditions (i)–(iii) above, you obey Reflection. So maybe we could justify Reflection by appealing to a normative principle that requires (i)–(iii).

The trouble with this approach is that no such principle can be correct. Certainly we oughtn't think that we will always conditionalize since we have overwhelming evidence that we often don't. Examples like the drinking case are enough to illustrate this, but there is evidence that we violate conditionalization more generally (see (Kahneman and Tversky 1982), for example). So (i) is not a normative requirement. The same goes for (iii). We clearly aren't luminous, so surely we ought not be certain that 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 4.1, that it can't be regarded as a norm of rationality.

3 OBJECTIONS

One might respond that, while (i)–(iii) are not normative requirements for us, they are normative requirements for a more ideal agent who is, in fact, a conditionalizer. If you are the sort of idealized Bayesian conditionalizer van Fraassen envisions, maybe it's plausible that you actually should be certain you will always conditionalize. And maybe you should be luminous too — certain that you will be luminous, even. Hence, the objections goes, the revised result shows that ideal agents, who are conditionalizers, ought to satisfy Reflection.

It's a bit unclear to me why an agent ideal enough to be a conditionalizer would be luminous, or have any reason to think that it would be luminous in the future. We can program a computer to be a perfect conditionalizer but have horrible introspective access; clearly there's no necessary connection. If the thought is just that a human who attained such Bayesian heights as to be a perfect Conditionalizer would be luminous, well, we don't have any reason to think that. But there's something more deeply wrong with this objection. Even if (i)–(iii) are satisfied at that level of idealization, the fact that agents at that level obey Reflection doesn't have any bearing on us. One can't reliably argue that, because ideal agents do a thing, we should too. After all, ideal agents never apologize because they never have to. But that's not a practice we should aspire to. Very plausibly this analogy is actually quite apt. Agents at the level of idealization under discussion have every reason to trust their future credences: they think they are perfect conditionalizers with perfect introspective access. But we are not so reliable, so we should not put too much stock in what we'll think in the future.

But, if I am right that Conditionalization does not imply Reflection, then I owe an explanation. One of van Fraassen's motivations for connecting Conditionalization to Reflection was to redirect all attacks on Reflection at proponents of Conditionalization. If counterexamples to Reflection were also counterexamples to Conditionalization, then we fans of Conditionalization would have to stop thinking of those counterexamples as counterexamples. I've argued that this isn't so. And yet, van Fraassen does seem to be on to something here. Recently, cases that appear to be rational violations of both Reflection *and Conditionalization* have been cropping up. For example, the infamous case of Sleeping Beauty (Elga 2000) is often thought to be one where Conditionalization and Reflection are rationally violated together. Arntzenius (2003) also offers some 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.

Take Arntzenius's Shangri-la case as an illustration. The deities have granted you a visit to Shangri-la, but they require that no one who comes to Shangri-la know how she 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 currently 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.

If Conditionalization doesn't entail Reflection, why is the violation of Reflection in this case accompanied by a violation of Conditionalization? Notice that the violation of Reflection happens en route while the violation of Conditionalization happens upon your arrival at Shangri-la. This is a cue that the two violations are not necessarily tied. In fact, we could have stopped the story at the sea and gotten our violation of Reflection without worrying about what happens next. Indeed, if we tell the story a bit differently so that, when you arrive, you do not drop your credence in heads to 1/2 but instead stick to 1, we have a violation of Reflection without violating Conditionalization. The reason the violation of Conditionalization does happen in Arntzenius's case is that, in the natural telling of the story, you actually do what you think you will do: namely violate Conditionalization on your arrival. Since you violated Reflection, the revised result from section 2 tells us you had to think you would violate Conditionalization all along. The natural way to tell the story is that you were right about that, so Conditionalization gets violated. But that's not the only way to tell the story. We can perfectly well imagine that you stick to your guns and remain certain of heads upon your arrival.

In general, 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 and Arntzenius, 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's and Elga's. You don't actually need to become uncertain about heads when you arrive at Shangri-la in order to to reasonably violate Reflection. All that's required is that you reasonably believe you will.

Now, none of this does anything to solve the challenge that these cases pose for Conditionalization. It just shows that this suspicious coincidence — violations of Reflection turning out to be violations of Conditionalization too — doesn't tell against my claim of logical independence. So what do I say about these cases qua violations of Conditionalization? I say they are not counterexamples to Conditionalization in the form I endorse it. My view is that we should conditionalize when we get new information.[10] An often accompanying view is that conditionalizing is the *only* way we should ever change our degrees of belief. It's this second claim that Elga's and Arntzenius's cases disprove. They show that, in cases where you lose information, it's reasonable to change your credences in a non-conditionalizing way. In the Shangri-la case, you lose information about which path you took. But this is something we had to accept anyway, just on the grounds that ordinary memory degradation is not irrational. It's a very interesting question how various types of information loss should be treated. But it does not challenge the modest claim that, in mundane cases where new information is learned and none is lost, Conditionalization is the way to go.

4 OTHER CONNECTIONS

I'm trying to make a case that Conditionalization and Reflection are thoroughly independent principles. We've seen that Conditionalization doesn't entail Reflection, but there's more work to be done. First, there's an argument that Conditionalization motivates Reflection with the help of some added assumptions, even if it doesn't entail it. In order to fully secure the possibility of Conditionalization without Reflection, I have to dispatch this argument. In

[10] Well actually, I don't think Conditionalization is exactly right for the usual reason: it requires us to give certainty to new evidence. But it's pretty close, so we can treat it as right for most purposes.

section 4.1 I show that it, like van Fraassen's argument, trades on implausible assumptions about introspective access.

There's also the converse question: does Reflection entail Conditionalization? Van Fraassen argues that it does, thereby offering a new justification of Conditionalization. This claim doesn't bear directly on my main aim of defending Conditionalization without Reflection, but it's a closely related and very interesting question. So, in section 4.2 I'll take a brief detour from my main theme to address it.

4.1 *The Principal Principle Analogy*

Superficially at least, there's a strong resemblance between Reflection and Lewis' Principal Principle (Lewis, 1980). 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 that 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[11] but not vice versa. So this way of connecting Conditional-

[11] 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

ization 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 credences before you get the evidence. If you obey Reflection and have $p(H|p_t(H)=x)=x$ right before t, then conditionalizing on $p_t(H)=x$ will make it true that $p_t(H)=x$. If you violate Reflection and have $p(H|p_t(H) = x) \neq x$ instead, conditionalizing on $p_t(H) = x$ will make it false that $p_t(H)=x$. 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.[12]

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?

these other y values is correct, i.e. $p_t(e)=x$.

[12] At the end of section 2 I promised an argument that Luminosity is not a norm of rationality. Since Luminosity implies Transparency, my critique of Transparency here is a critique of Luminosity as well.

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 believe in leprechauns, Reflection will require that I adopt that belief 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 leprechauns, 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 the general problem with arguing from ideals to norms that I mentioned in section 3. 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 should 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 (recall: ideal agents never apologize because they never have to). Of the possible outcomes that are attainable for us, X-outcomes may be less than ideal.

4.2 *Van Fraassen's Argument for the Converse*

So much for deriving Reflection from 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.

Roughly speaking, GR is supposed to require Conditionalization when the agent is certain that her evidence at some future time will be one of the elements of a partition. We don't need to worry about how this assumption gets used to prove the result van Fraassen derives from it, because its the significance of the result that I'm going to dispute. To state the result, let's let q_i be the distribution the agent thinks she will come to have if she receives E_i as evidence. Van Fraassen shows

Result If the agent satisfies GR then $p(\cdot|E_i) = q_i$ for every i such that $p(E_i) > 0$.

This is supposed to show that an agent who satisfies GR 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, however, the problem here should be fairly apparent. While it's trivially 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 opinions.

5 LIFE WITHOUT REFLECTION

We've seen that Conditionalization and Reflection are logically independent; neither entails the other. But just because there's room in logical space for Conditionalization without Reflection doesn't mean that we've settled the matter. In some cases, Reflection-type constraints look quite rational, and this needs to be accounted for. To make life without Reflection as comfortable as possible, I want to say something positive about how to account for those judgments without Reflection.

Let's start by taking a case where Reflection-type reasoning looks good. Suppose I'm not sure whether Oswald really shot Kennedy, so I ask a knowledgeable friend to recommend a book on the subject. She lends me a recent book arguing that Oswald did in fact shoot Kennedy, saying that it's the best work done on the topic; the author presents heaps of good evidence for his thesis, and thoroughly debunks the evidence of his opponents. As I sit down with the book I realize that, when I finish it, I'll be much more confident that Oswald shot Kennedy. So shouldn't I be just as confident now? It seems clear to me that I should.

So there are some cases where obeying Reflection is correct and some cases where it isn't (recall the drinking case). The challenge is to say what distinguishes the good cases, and how we should account for them without appealing to Reflection. Looking at the Oswald book example, what distinguishes it from the drinking case seems to be that your future credence will result from you learning lots of strong evidence. This points to an attractive general criterion for distinguishing good cases from bad ones: trust your future credences when they will derive from good evidence, and not otherwise. But making this criterion precise is tricky. If your future self is going to forget some of the things you know now, she may interpret evidence very differently from the way you would. In that case, you wouldn't want to trust her credences. For example, if she has forgotten that red litmus paper means acid and blue means base, you shouldn't adopt whatever conclusion she draws from a litmus paper test. Also, even if your future self hasn't forgotten anything and has learned lots of new evidence, she might react to it in a way that you would consider irrational. If your future self is drunk, or inclined to project grue instead of green, you don't want to trust her then either.

To get rid of these problems let's clean up our criterion: trust your future credences when they will be based on strictly more evidence than you yourself have now, and will be derived from that evidence in a way you deem rational. If your future self knows more and draws conclusions just as you see fit, you should take her opinion very seriously. This takes care of the worries just raised, and suggests replacing Reflection with the following:

Trust Your credence in H, given that your future self would have credence x in H if she (i) had all your evidence in addition to hers, and (ii) evaluated her evidence rationally (by your lights), should be x.

Condition (i) rules out cases where your future self isn't privy to information that you've got, like the litmus example, and condition (ii) rules out cases where your future self reacts to evidence irrationally, like in the drinking case.

It's tempting to formalize Trust as something like $p(H|p_t(H|\text{total evidence})=x)=x$. This expresses basically the same idea, except that the condition that your future self react to evidence rationally (by your lights) is being replaced by the condition that she conditionalize on it. In fact, Elga (unpublished) endorses something similar. But I think this formalization isn't exactly right. What if you don't know that Conditionalization is the right way to weigh evidence? If you think that some other rule is the right way to update your credences, you shouldn't put your faith in your future conditional probabilities. At least, it's far from clear to me that you should. After all, if you think Conditionalization is irrational and yet you adopt your future conditional probabilities, you are in effect saying "I think Conditionalization is an irrational policy but I am going to follow it anyway." But even if you are sure that Conditionalization is rational, there's the worry that your future self will have undergone a radically irrational change in probabilities, so that her conditional probabilities don't look anything like yours. If she has decided to project grue instead of green, her conditional probabilities don't amount to weighing evidence in the way that you would. So you shouldn't trust them. For these reasons, I think we should stick with Trust as is.

Trust does a better job of capturing the idea behind Reflection than Reflection itself. Trust, I claim, accurately captures correct applications of Reflection-type reasoning without leading to Reflection's bad consequences. In the Owsald book example, you learn that you will have a high credence in Oswald's guilt that is rationally based on strictly more evidence than you have now. If you satisfy Trust, conditionalizing on that information will give you that high credence, as seems right. In the drinking case, however, you know your future credences will not be rationally based on your evidence, so Trust doesn't apply (though Reflection would). Trust and Conditionalization in hand, life without Reflection is good.

REFERENCES

Arntzenius, Frank. 2003. "Some Problems for Conditionalization and Reflection." *Journal of Philosophy* C.7.

Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press.

Elga, Adam. 2000. "Self-Locating Belief and the Sleeping Beauty Problem." *Analysis* 60.

Elga, Adam. N.d. "Solidarity and Disagreement." unpublished.

Kahneman, Daniel and Amos Tversky. 1982. Judgments of and by Representativeness. In *Judgment Under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic and Amos Tversky. Cambridge University Press.

Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance." *Studies in Inductive Logic and Probability,* II.

Schick, Frederick. 1986. “Dutch Bookies and Money Pumps.” *Journal of Philosophy* 83.

Talbott, William J. 1991. “Two Principles of Bayesian Epistemology.” *Philosophical Studies* 62.

van Fraassen, Bas. 1984. “Belief and the Will.” *The Journal of Philosophy* 81.

van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford University Press.

van Fraassen, Bas. 1995. “Belief and the Problem of Ulysses and the Sirens.” *Philosophical Studies* 77.

van Fraassen, Bas. 1999. “Conditionalization: a New Argument For.” *Topoi* 18.

Weatherson, Brian. 1999. “Begging the Question and Bayesians.” *Studies in the History and Philosophy of Science* 30.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.