

W. J. TALBOTT

TWO PRINCIPLES OF BAYESIAN EPISTEMOLOGY*

(Received 1 March, 1990)

Van Fraassen (1984) has argued that what he calls the *Principle of Reflection* is a requirement of epistemic rationality: Where $p_t^a(A)$ is agent a 's assignment of subjective probability to (or degree of credence or degree of belief in) proposition A at time t , $0 \leq r \leq 1$, and $x \geq 0$:

$$\text{(Reflection)} \quad p_t^a(A | p_{t+x}^a(A) = r) = r^1$$

Reflection says that at time t , given that a 's probability assignment to A at $t + x$ is (or will be) r , a 's current probability assignment to A is (or should be) r .

It is useful to decompose Reflection into two parts: The first, uncontroversial part I refer to as the Principle of *Concurrent Reflection*. Concurrent Reflection is the limitation of van Fraassen's Reflection Principle to cases in which $x = 0$.² The second, more controversial part is the Principle of *Future Reflection*. Future Reflection is the limitation of van Fraassen's Reflection Principle to cases in which $x > 0$.

Since the article in which he first stated and defended his view that Reflection is a necessary condition for epistemic rationality, van Fraassen has apparently qualified his view to the extent that he is now willing to allow for states of less than total rationality (e.g., being as rational as possible under the circumstances), in which Future Reflection is violated.³ But he continues to defend Future Reflection as a requirement of ideal or total rationality.

In Section I below, I introduce an example of the kind which led van Fraassen to qualify his endorsement of Future Reflection — that is, an example involving a *prima facie* irrational future belief (or degree of belief). I argue that such examples need not cast doubt on Future Reflection, if Future Reflection is understood to operate jointly with the well-known Bayesian principle of Temporal Conditionalization.

However, in Section II below, I show that violations of the joint

principles of Temporal Conditionalization and Future Reflection are much more widespread than might be supposed, and that they include cases in which there is no *prima facie* irrationality on the part of the agent. This leads me to reconsider the arguments — including the diachronic Dutch Book arguments — that Future Reflection and Temporal Conditionalization are requirements of epistemic rationality. Ultimately, I argue that these two principles of Bayesian epistemology, Future Reflection and Temporal Conditionalization, are very different from the other normative conditions in the literature, and that they are neither individually nor jointly necessary for epistemic rationality.

I. FUTURE REFLECTION AND TEMPORAL CONDITIONALIZATION

Consider an agent a who is about to leave for a party at which she knows she is likely to drink too much. It may be rational for a to make arrangements to ride with a friend who does not drink, and to intentionally leave her own car at home, if she knows that when she is under the influence of alcohol she tends to underestimate its effects on her mental and physical functioning.⁴ In this situation, where the irrationality of her future epistemic state is a given, to be rational (at least as rational as she can be under the circumstances) seems to require that the agent a violate Future Reflection:

Let J be the proposition that at $t + x$ (i.e., at the end of the evening) the agent a 's judgment and reactions will be — or (tenselessly) are — significantly impaired. At t (i.e., at the beginning of the evening, before she has done any drinking), it may be rational for agent a to be quite sure of J , while acknowledging that at the end of the evening she will vociferously deny it — that is, both (1) and (2) hold at t :

$$(1) \quad p_t^a(J) \approx 1$$

$$(2) \quad p_t^a(p_{t+x}^a(J) \approx 0) \approx 1$$

Future Reflection requires:

$$(3) \quad p_t^a(J | p_{t+x}^a(J) \approx 0) \approx 0$$

Conjoined with (2), Future Reflection requires that, at the beginning of

the evening (before she has done any drinking), the agent have the same confidence in the falsity of J that she knows she will have at the end of the evening (after imbibing freely). Formally, (2) and (3) entail (4):

$$(4) \quad p_i^a(J) \approx 0$$

And (4) contradicts (1).

In this case, it is clear that at t (before she has done any drinking), it would be irrational for a to replace (1) with (4). Given what a knows about the effect of drinking on her faculties, replacing (1) with (4) would have the effect of importing a 's irrationality at $t + x$ into what we can assume to be a 's otherwise rational probability assignment at t . Thus, to enforce Future Reflection in this case would only multiply a 's irrationality. In this case, rationality — or, at least, being as rational as one can be under the circumstances — requires a to favor (1) over (4), and thus to violate Future Reflection at t .

An advocate of Future Reflection can accommodate examples such as this one by limiting the applicability of Future Reflection to situations in which the agent a knows (or believes) that her probability assignment (to the relevant propositions) at $t + x$ is rational. To avoid circularity, an advocate of Future Reflection could argue for a condition on the applicability of Future Reflection as follows:

The principles of epistemic rationality must be understood to be *joint* conditions on rationality. In the foregoing example, Future Reflection is inapplicable, because the agent knows that she will violate a separate and joint requirement on epistemic rationality, *Temporal Conditionalization*: Where $p_t^a(A)$ is agent a 's probability assignment to A at t and E is a proposition stating all the new evidence that a acquires between t and $t + 1$, then $p_{t+1}^a(A)$, a 's probability assignment to A at $t + 1$, will not be rational unless:

$$(\text{Temporal Conditionalization}) \quad p_{t+1}^a(A) = p_t^a(A|E)^5$$

If Temporal Conditionalization and Future Reflection are understood to be joint conditions of epistemic rationality, then the rationality of violating Future Reflection in the above example can be explained by the agent's realization that she will violate Temporal Conditionalization: From a 's vantage point at time t (before she goes out drinking), she realizes that during the course of the evening she will consume a

substantial amount of alcohol and that she will obtain plenty of evidence that her judgment and reactions are very much influenced by the alcohol that she has consumed. Were it not for the effects of the alcohol, that evidence would (via Temporal Conditionalization) lead her by the end of the evening (at time $t + x$) to assign a high probability to the proposition J (that her judgment and reactions have been impaired by the alcohol). But the alcohol will prevent her from properly conditionalizing on the evidence, and as a result, at the end of the evening (i.e., at time $t + x$) her degree of belief in J will be irrationally low.⁶

Given agent a 's knowledge at t that her degree of belief in J at $t + x$ will not be in accordance with the requirements of Temporal Conditionalization, and therefore not be rational, for agent a to satisfy Future Reflection (with respect to her irrational degree of belief in A at $t + x$), would only compound a 's irrationality at $t + x$, by importing it into what we can assume to be a 's otherwise rational probability assignment at t . Thus, given agent a 's knowledge at time t that at $t + x$ she will violate one of the joint requirements of rationality (Temporal Conditionalization), rationality leads her, at time t , to violate another of the joint requirements (Future Reflection) also.

As applied to opinions influenced by alcohol and other *prima facie* irrational factors, an account of this kind seems to fit the data well. The problem for any such account is that there are other cases — cases in which we are not at all inclined to ascribe any irrationality to the agent — which seem to be indistinguishable from this drinking example and the other examples involving *prima facie* irrationality. That is to say, there are examples in which, being as rational as it is humanly possible to be, we unavoidably violate Temporal Conditionalization, and our recognition that we will violate Temporal Conditionalization leads us to violate Future Reflection also.

II. ANTICIPATING FUTURE MEMORY LOSS

Consider the following example: I can remember what I had for dinner one week ago, but I have no memory of what I had one month ago or one year ago. This is not usually taken to reflect on my rationality, for there can be little doubt that if philosophers held a special election to

name the most epistemically rational human being alive, neither the winner of that election nor any of the top finishers would, except on certain special occasions, remember what they had eaten for dinner one year earlier either.

If we assume that I had spaghetti for dinner one year ago (on March 15, 1989), then where t is 6:30 p.m. on March 15, 1989; $t + x$ is 6:30 p.m. on March 15, 1990; and S is the proposition that Talbott had spaghetti for dinner on March 15, 1989; it is clear that at time t (March 15, 1989), rationality would have required me to violate Future Reflection, because I knew that, though not due to any irrationality on my part, I would violate Temporal Conditionalization at $t + x$ (March 15, 1990) — that is:

At t I knew that I was eating spaghetti for dinner. Thus, in light of the simplifying assumption that I assign subjective probability one to my beliefs (fn. 1 above), where $T =$ Talbott:

$$(5) \quad p_t^T(S) = 1$$

During the year from t to $t + x$, I have learned nothing with any evidential bearing on what I had for dinner at t — that is, if $\varepsilon = \{E_1, \dots, E_n\}$ is the set of propositions stating all the evidence that I have acquired in the period from t to $t + x$, every member E_i of ε is such that:

$$(6) \quad p_t^T(S|E_i) = p_t^T(S)$$

Therefore, if I am assumed to satisfy Temporal Conditionalization at every time from t to $t + x$, it must be the case that I be as certain of S at $t + x$ as I was at t — that is, where $\&E_i$ is the conjunction of all of the members of ε , from (5) and (6):

$$(7) \quad p_{t+x}^T(S) = p_t^T(S|\&E_i) = 1$$

But, in fact (and I knew this at t), if at $t + x$ I am asked how probable it is that I had spaghetti for dinner on March 15, 1989, the best I can do is to calculate a probability based on the relative frequency of spaghetti dinners in my diet one year ago. Doing so, I calculate the probability of S :

$$(8) \quad p_{t+x}^T(S) = 0.1$$

Because I know (8) at t:

$$(9) \quad p_t^T(p_{t+x}^T(S) = 0.1) = 1$$

However, Future Reflection requires that at t:

$$(10) \quad p_t^T(S|p_{t+x}^T(S) = 0.1) = 0.1$$

Conjoined with (9), Future Reflection requires that I be as much in doubt about what I am eating for dinner while I am eating it as I know that I will be one year later. Formally, (9) and (10) imply:

$$(11) \quad p_t^T(S) = 0.1$$

Of course, (11) contradicts (5). Thus, if I am to maintain my near certainty at t that I am eating spaghetti for dinner [(5) above], my knowledge at t that I will violate Temporal Conditionalization at t + x will lead me to violate Future Reflection at t.

Formally, this example is equivalent to the drinking example discussed earlier. In that case the failure of Future Reflection (at t) was explained by the prima facie irrationality of the agent's probability assignment at t + x, where that irrationality was further explained by a failure of Temporal Conditionalization. But in this case, though there is an evident failure of Future Reflection (at t) based on the agent's awareness that there will be a failure of Temporal Conditionalization at t + x, there is no prima facie irrationality involved.

Where my failing to remember what I had for dinner one year ago is not due to self-deception or any bad faith on my part, and where there are no other prima facie irrational influences (e.g., alcohol) involved, it seems that at both t and t + x (i.e., in holding both (5) and (9)) above, I am being as rational as it is humanly possible to be in my probability assignment to S. Indeed, where I have made no record of what I had for dinner one year ago, and there is no other source of the information, and thus where it would be impossible for me by any rational procedure to acquire the degree of confidence in S that Temporal Conditionalization would require of me (given by (7) above), the advocate of these two principles of Bayesian epistemology, Temporal Conditionalization and Future Reflection, seems to be committed to a standard of rationality that, at least in these cases, there is no rational way for me to satisfy.

To defend the two Bayesian principles, one could argue that there is no important difference between the drinking example and the example involving memory decay. Both examples would be analyzed as departures from ideal rationality. An ideally rational agent would not suffer impairments of her faculties and judgment, and an ideally rational agent would not be subject to the fallibility of human memory.

However, by conflating the analysis of the two cases, this response would not be able to account for the strong sense that the beliefs influenced by alcohol are *irrational* in a way that beliefs resulting from memory decay are not. And even if that intuition is simply mistaken, it must be admitted that, at the very least, the beliefs influenced by alcohol are *unreliable* in a way that the beliefs resulting from memory decay are not — for example, if my probability assignment of 0.1 to *S* at $t + x$ is based on the relative frequency of spaghetti dinners in my diet, it should be quite reliable in the long run. But the two Bayesian principles cannot account for this difference either.

Finally, this response opens up such a great distance between ideally rational agents and human beings that we would have no realistic prospects of ever even approximating the ideal. It would leave us with no standard by which to measure the rationality of the vast preponderance of our beliefs (and degrees of belief) that are based, at least in part, on fallible, human memory.

Not only would the rationality of most opinions about the not-just-recent past be left a mystery, but the rationality of many theoretical opinions also; for no matter how they were originally acquired — e.g., in conversation or by working out a derivation of them from already accepted premises — one's continuing belief (or degree of belief) in them will be based on memory.

An "ideally rational" agent who satisfied Temporal Conditionalization and Future Reflection would not have forgotten any of the information she had acquired in the past, and, in the absence of contrary evidence, would not have any reason to reduce her confidence in beliefs acquired in the past from the level of confidence she had in them at the time she acquired them. For such an agent, it would not be rational to reread an article before referring to it in a lecture or published work, or to reread her own work to be sure she had said what she intended to say. But for a human being, not to do so would

often be disastrous. Because such extreme self-confidence would be positively required of a rational agent who satisfied Temporal Conditionalization and Future Reflection, its irrationality for human beings cannot be explained by reference to such “ideally rational” agents.

III. THE DISANALOGY WITH OTHER IDEALIZING ASSUMPTIONS

In the Bayesian tradition, works on epistemic rationality typically assume that a rational agent's beliefs (or degrees of belief) are *deductively closed* and that the agent's degrees of belief are *synchronically coherent* (i.e., that they satisfy the probability axioms). These assumptions are extreme idealizations, the significance of which should not be underestimated.⁷ However, it should be emphasized that one of the common defenses of these idealizations is not available to the defender of Temporal Conditionalization and Future Reflection.

Levi (1980, pp. 10–11), for example, defends deductive closure by distinguishing an agent's epistemological *commitment* at *t* from her *awareness* of her commitment at *t*. For Levi, deductive closure can be used to define the agent's commitments, even if the agent herself is not aware of all of her commitments. This distinction is defensible, because, in Levi's terms, the agent's corpus of knowledge contains all of the items (premises and rules of inference) needed to derive her “commitments”, and thus, given a derivation of a proposition *T* from premises S_1, \dots, S_n in the agent's corpus, in accordance with axioms and rules of inference in the agent's corpus, the agent can be expected to acknowledge her commitment and accept *T* into her corpus.⁸

Moreover, it is at least not implausible to think that there might be generally acceptable principles which, when applied to a synchronically incoherent subjective probability assignment, would transform it into a synchronically coherent assignment. If so, then it could be said that the person with the incoherent probability assignment was committed to the coherent one, in very much the same sense that a person can be said to be committed to the logical consequences of her beliefs.

The contrast with memory effects is evident. Although Temporal Conditionalization would require me to be quite sure of what I had had for dinner one year ago (which, for the sake of the present discussion,

we have assumed to be spaghetti), where I have not the slightest memory of that dinner and no other evidence of what I had, there is no plausible sense in which my current beliefs (or degrees of belief) could be said to contain a commitment to (assigning high probability to) the proposition that I had spaghetti for dinner one year ago. No rational reconstruction of my current beliefs (or degrees of belief) by any plausible rules of deductive or inductive inference will imply (my assigning high probability to) the proposition that I had spaghetti for dinner one year ago, because there is no trace of a memory of that dinner left. The only way I could ever discover what I had for dinner one year ago would be to acquire new evidence concerning it. Therefore, even someone who favors incorporating the deductive closure or synchronic coherence requirements into the theory of ideal rationality is in no way committed to the very different kind of “idealization” that results from adding Temporal Conditionalization and Future Reflection.

IV. THE DUTCH STRATEGY ARGUMENT

There is one respect in which Temporal Conditionalization and Future Reflection are analogous to the synchronic coherence conditions. All have been defended by Dutch Book arguments — synchronic coherence by a synchronic Dutch Book argument, Temporal Conditionalization and Future Reflection by a diachronic Dutch Book argument or as van Fraassen (1984) calls it, a “Dutch Strategy”.⁹ A synchronic Dutch Book is a group of wagers offered at a single time t , each of which is judged to be fair by the agent at t , and which, in the aggregate, make the agent a sure loser. A diachronic Dutch Book (or Dutch Strategy) is a sequence of wagers at t and $t + x$ such that each of the wagers offered at t is judged to be fair by the agent at t , and each of the wagers offered at $t + x$ is judged by the agent to be fair at $t + x$, where the combination of the bets at t and $t + x$ makes the agent a sure loser.

Susceptibility to a synchronic or diachronic Dutch Book combination of wagers may seem to be irrational. However, whether it is irrational depends crucially on whether the agent herself is aware, or ought to be aware, that the combination is a Dutch Book. Because no human being is logically omniscient, any human being, no matter how epistemically rational, will be susceptible to a potential Dutch Book. Lacking logical

omniscience, one would simply be unaware that the relevant wagers comprised a Dutch Book.

Though the existence of a potential Dutch Book cannot by itself show that an agent is epistemically irrational, subject to one qualification that I take up below, it does seem to be a minimal condition of epistemic rationality that an agent not voluntarily accept an *explicit* Dutch Book — that is, a sequence of wagers (whether synchronic or diachronic) which, at each stage, the agent recognizes to be part of a Dutch Book combination.¹⁰ In light of the foregoing discussion of the epistemic role of memory, which, if successful, shows that human beings, no matter how rational, do not and cannot and ought not to satisfy Temporal Conditionalization and Future Reflection, must one conclude that even the most rational human beings will be susceptible to explicit Dutch Strategies?

No. Consider again the spaghetti dinner example. It is true that, if, as I have supposed, at t (i.e., one year ago, on March 15, 1989), I was eating spaghetti for dinner, then at that time I would have been willing to give good odds on S (the proposition that Talbott had spaghetti for dinner on March 15, 1989). Ignoring inflation and other discounts, I would have paid at least 95 cents for a dollar of return. And it is also true that at $t + x$ (i.e., one year later, on March 15, 1990) in my current state of ignorance about what I had for dinner one year ago, I would sell a wager paying one dollar if S is true for about 10 cents. But it does not follow that I would have been susceptible to an explicit Dutch Strategy.

For example, if at time t , when entering into the original wager on S , I had been aware that I would also be offered a wager on $\neg S$ one year later, it might well have been rational for me to make a record of my dinner — if the stakes were high enough, to make a videotape of it. Then one year later, at time $t + x$, I could consult the videotape, find out that I had spaghetti for dinner at t , and refuse the second half of the potential Dutch Strategy combination of wagers.

Moreover, even if at $t + x$ I had no record of what I had eaten for dinner at t (perhaps I made a videotape of my dinner at t , but now, at $t + x$, I cannot remember where I stored it), I would still not be susceptible to an *explicit* Dutch Strategy. For the Dutch Strategy to be explicit, at $t + x$ I would have to realize that the proffered wager at

$t + x$ was the second half of a Dutch Strategy combination, the first half of which was an earlier wager, giving high odds, on S at t . But if at $t + x$ I knew that I had offered high odds on S at t (while eating the dinner described by S), that additional information would change my probability assignment to S at $t + x$. In the absence of that information, the best I could do at $t + x$ would be to assign a probability to S equal to my estimate of the relative frequency of spaghetti dinners in my diet one year earlier (which I have assumed to be .1). But given the additional information that I was almost certain of S at the time that I was eating the dinner described by S , it would be rational for me to be as certain of S at $t + x$ as I was at t . And thus, in the spaghetti dinner example, I would not be susceptible to an explicit Dutch Strategy.

Finally, as a practical matter, it may even be rational to acquiesce in a Dutch Strategy, when doing so is part of a larger plan with an expected gain. Thus, for example, suppose that I have a standing even-money bet with the Dutch bookie that I did not have spaghetti for dinner one year earlier. We make the same bet every day. I win nine bets for every one that she wins. Today is March 15, 1990. Although I do not yet know it, today I am going to lose our standing bet. (I had spaghetti for dinner one year ago, although I don't now remember that I did.) When I lose the standing bet, I will have completed the second part of the Dutch Strategy that made me a guaranteed loser on the combination of my March 15, 1989 bet that S is true and my March 15, 1990 bet that S is false. It should be evident that, in this context, my susceptibility to the Dutch Strategy is no threat to my wealth. Because I expect to win all the bets made at the time that I am eating spaghetti for dinner and 9/10s of the bets on what I had for dinner one year earlier, I can expect a large net gain from the entire series of wagers, though in some instances I will suffer what in isolation would seem to be a sure loss. Indeed, when acquiescing in the Dutch Strategy is seen to be part of an overall strategy with a positive return, it may be rational for me to do so, even if I know that I am doing so. For example, even if I know that I had spaghetti for dinner one year ago, and therefore that I will lose my bet on $\neg S$ today, it may be rational for me to bet on $\neg S$ today, if my doing so is a condition of the bookie's continuing to offer the bet in the future. Thus, examples of the failure of Temporal Conditionalization and Future Reflection due to the fallibility of memory either do not

entail susceptibility to an actual Dutch Strategy, or the Dutch Strategy turns out to be harmless.

It may seem that the foregoing argument undercuts my earlier strictures against Temporal Conditionalization and Future Reflection, because my advice for avoiding the Dutch Strategy — e.g., in the case where I suggested making a videotape — amounted to advising the agent to take steps to insure that Temporal Conditionalization and Future Reflection were not violated. So isn't my argument really a veiled defense of these two principles of Bayesian epistemology?

Not at all. My response to the Dutch Strategy arguments only implies that one should take steps to satisfy Temporal Conditionalization and Future Reflection in those relatively rare instances when it pays to do so. Any theory which assumed Temporal Conditionalization and Future Reflection as constitutive of epistemic rationality could not even begin to identify such instances.

I have not, for example, argued that one should videotape one's entire life, so as to have a complete record of all one's evidence. Although there are undoubtedly cases in which it would be useful to have such videotapes — for example, to settle an argument about who said what on some occasion in the past — only someone in very unusual circumstances or with very unusual preferences would find it rational to make the necessary videotapes. And even if such videotapes existed, it would often be rational to violate Temporal Conditionalization rather than taking the time to review them. For example, even if there were complete videotapes that I could consult to resolve a disagreement between my two daughters on the relative frequency of spaghetti dinners in our diet in the past year — one daughter claiming that the relative frequency is greater than 0.1 and the other that it is less than 0.1 — it would be irrational for me to spend the hours reviewing them that would be necessary to determine the actual relative frequency of spaghetti dinners in our diet and thus to resolve the dispute. No acceptable theory of epistemic rationality will require me to spend many hours reviewing videotapes merely to avoid violating Temporal Conditionalization with respect to such an inconsequential question.

Making videotapes of one's entire life is one strategy for attempting to satisfy Temporal Conditionalization and Future Reflection. Another strategy would be to limit one's evidence to such a small number of

items that they could all be easily remembered. For example, one might live as a recluse, on a strict routine of exactly the same activities every day. If one always ate the same meals at the same times, there would be no difficulty in satisfying Future Reflection and Temporal Conditionalization with respect to propositions about one's past and future dinners. Again, however, only someone in very unusual circumstances or with very unusual preferences would find it rational to live such a life. Any theory that implies that one should live one's life in such a way as to learn as little as possible should have a difficult time gaining much credence as a theory of epistemic rationality.

V. CONCLUSION

There are many cases in which an epistemically rational human being will realize that, through no irrationality on her part, she will fail to satisfy Temporal Conditionalization, and in which this realization will lead her to violate Future Reflection also. Violations of these two principles of Bayesian epistemology are distinguishable from other cases in which human beings fail to satisfy an epistemic ideal (e.g., failures of deductive closure or of synchronic coherence), because, in the cases of interest, there is no plausible sense in which the agent who violates the two Bayesian principles can be thought to be *committed* to probability assignments that do not violate them. Thus, whether taken individually or jointly, these two principles of Bayesian epistemology, Temporal Conditionalization and Future Reflection, are not necessary conditions for any kind of epistemic rationality that human beings can achieve or approximate, or rationally ought to attempt to achieve or approximate.

NOTES

* An earlier version of this paper was presented at the Colloquium on Logic, Probability, and Methodology at the A. P. A. Eastern Division Meetings in December 1987. I am indebted to all those who participated in that discussion, and especially to Teddy Seidenfeld and Bas C. van Fraassen.

¹ Following the usual practice, I assume that the agent's subjective probability assignment satisfies the probability axioms, but my positive claims do not depend on that assumption, and can be interpreted to apply to subjective confidence assignments that are not completely coherent, in the usual sense. I also employ the usual definition of

conditional subjective probability of A given B : $p(A|B) = p(A \& B)/p(B)$ when $p(B) > 0$, and is undefined if $p(B) = 0$.

Van Fraassen (1984) uses both upper case and lower case letters ("P" and "p") to refer to an agent's subjective probability assignment. Because van Fraassen's "P/p" distinction will not come into play in this paper, I have used the lower case "p" throughout.

In addition to a subjective probability assignment, it will simplify the subsequent discussion to assume that an epistemic agent a has a set $\{B_i\}_t^a$ of beliefs at t and also to assume that $\{B_i\}_t^a$ contains only propositions of which a is certain at t . Thus, where B is any member of (or conjunction of members of) $\{B_i\}_t^a$, a 's set of beliefs at t , and S is any proposition to which a assigns subjective probability at t : $p_t^a(S) = p_t^a(S|B)$.

Finally, it should be noted that the expressions employed in stating the Principle of Reflection and the other principles employed in this paper are to be interpreted intensionally, not extensionally, and that truth is not preserved by substitution of co-extensive expressions, nor even by substitution of logically equivalent expressions. Similar strictures apply to my references to propositions and properties, and to statements of an agent's beliefs and statements of an agent's subjective probability assignments.

² Van Fraassen (1984) refers to the case where $x = 0$ as the *synchronic* part of Reflection (p. 248). I slightly favor *Concurrent Reflection* over van Fraassen's formulation, because there is a sense in which Concurrent and Future Reflection are *both* synchronic principles — that is, that they both are either satisfied or violated by the agent's subjective probability assignment at t (including her opinions at t about her subjective probability assignment at $t + x$). Future Reflection does not, by itself, impose any constraints on the relation between the agent's subjective probability assignment at t and the agent's subjective probability assignment at $t + x$.

³ For van Fraassen's more recent views, I am relying on an unpublished paper, "Belief and the Problem of Ulysses and the Sirens", and on personal discussions.

⁴ To avoid unnecessary complications, I assume here that the decision to imbibe is a rational one.

⁵ The Temporal Conditionalization Principle in the text is the traditional principle, which Teller (1976) refers to as *strict conditionalization*. Because strict conditionalization implausibly requires that the agent be certain about the evidence, it might best be understood as a special case of Jeffrey's (1965) conditionalization principle, which Teller calls *general conditionalization*. On Jeffrey's account, acquiring evidence is not taken to imply an *acceptance* of evidentiary propositions, but only an increased probability of their being true. To introduce Jeffrey's general principle into the current discussion would only add unnecessary complications. But it is important to note that the argument in the text is not limited to the strict principle. It applies to Jeffrey's general principle also.

Although the argument in the text is addressed to views that imply that all rational changes in degree of belief satisfy a Temporal Conditionalization principle, it can also be modified to apply to views, such as Levi (1980), that accord Temporal Conditionalization a more limited role. In addition to changes in credence satisfying what Levi calls *temporal credal conditionalization* — Levi's analogue of Temporal Conditionalization, where, as in the examples discussed in the text, rigidity can be assumed to hold (that is, where it can be assumed that there is no change in the relevant conditional probabilities from t to $t + 1$) — Levi allows for contraction of the corpus of knowledge (i.e., the removal of propositions that were taken as certain before their removal). The examples discussed below raise difficulties for Levi's view also, because they do not fit either of Levi's (1980) criteria for legitimate contraction — that is, they are not situations where there is a need to remove a contradiction or to give a hypothesis a hearing (p. 61).

Finally, it should be noted that the argument does not depend on one's accepting

Temporal Conditionalization as a principle of epistemic rationality. I introduce Temporal Conditionalization because it is such an obvious candidate for the role of the relevant joint condition on epistemic rationality.

⁶ This is not the only possible explanation of how the alcohol might affect agent *a*'s probability assignment. In theory, it could also exert its effect by making her unable to make the observations that she would make if she were sober, or it could change her assignment of prior probabilities. In the text I have singled out the alternative that will be most relevant in the subsequent discussion.

⁷ Many philosophers have objected to deductive closure, most visibly Harman (1970) and Kyburg (1970). Garber (1983) has described one of the problems with the deductive closure assumption in confirmation theory, and has shown how, within a Bayesian framework, it can be weakened. Goldman (1986) objects to both deductive closure and synchronic coherence (pp. 311–318).

⁸ Levi (1980), pp. 6–7, 9–13. However, it should be added, as Harman (1970), has pointed out, that another alternative would be for the agent to discard one of the premises from which *T* was inferred. I do not wish to defend the deductive closure and synchronic coherence conditions. My point is just that even someone who believes that there is good reason to accept those principles should acknowledge that there is less reason to accept Temporal Conditionalization and Future Reflection.

⁹ The original Dutch Strategy argument is David Lewis's argument for Temporal Conditionalization, which was reported by Teller (1976, pp. 209–212). Van Fraassen (1984) has modified the Lewis argument into an argument for Future Reflection. It should be noted that, even on its own terms, the Lewis argument is not exactly an argument for Temporal Conditionalization, for any rule with the same expectation as Temporal Conditionalization will also thwart the proffered strategy — including many rules that are incompatible with Temporal Conditionalization. For example, where *E* is a proposition stating agent *a*'s total evidence acquired between *t* and *t* + *x* and where *d* = min(1 - $p_t^a(A|E)$, $p_x^a(A|E) - 0$), the following rule has the same expectation as Temporal Conditionalization: At *t* + *x* toss a fair coin; if it lands heads: $p_{t+x}^a(A) = p_x^a(A|E) + d$; if it lands tails: $p_{t+x}^a(A) = p_x^a(A|E) - d$. An agent who follows this rule is guaranteed not to satisfy Temporal Conditionalization (in all cases in which *d* > 0), and yet she will not be susceptible to Lewis's Dutch Strategy.

¹⁰ Strictly speaking, no epistemological theory by itself will imply anything about an agent's voluntary behavior. Thus, strictly speaking, what an epistemological theory should not imply is the rationality of an epistemic state that would incline an agent to voluntarily accept what she recognized to be a series of wagers on which she was sure to come out a loser (where it is assumed that there is no utility to the wagers other than the utility of their payoffs, as there would be, for example, if the agent wanted to give money to the bookie, but knew that she would refuse charity.)

REFERENCES

- Garber, Daniel (1983), 'Old Evidence and Logical Omniscience in Bayesian Confirmation Theory', in J. Earman (ed.), *Minnesota Studies in the Philosophy of Science*, Vol. X (Minneapolis: Univ. of Minn. Press), pp. 99–131.
- Goldman, Alvin I. (1986), *Epistemology and Cognition* (Cambridge: Harvard Univ. Press).
- Harman, Gilbert (1970), 'Induction', in M. Swain (ed.), *Induction, Acceptance, and Rational Belief* (Dordrecht: D. Reidel), pp. 83–99.
- Jeffrey, Richard C. (1965), *The Logic of Decision* (New York: McGraw-Hill).

- Kyburg, Henry E., Jr. (1970), 'Conjunctivitis', in M. Swain (ed.), *Induction, Acceptance, and Rational Belief* (Dordrecht: D. Reidel), pp. 55–82.
- Levi, Isaac (1980), *The Enterprise of Knowledge* (Cambridge, Mass.: MIT Press).
- Teller, Paul (1976), 'Conditionalization, Observation, and Change of Preference', in W. L. Harper and C. L. Hooker, (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume I*, (Dordrecht: D. Reidel), pp. 205–253.
- van Fraassen, Bas C. (1984), 'Belief and the Will', *Journal of Philosophy*, LXXXI, 5, pp. 235–256.
- van Fraassen, Bas C. (unpublished), 'Belief and the Problem of Ulysses and the Sirens'.

Department of Philosophy
University of Washington
345 Savery Hall
Seattle, WA 98195
U.S.A.