Diachronic Rationality
Author(s): Patrick Maher
Source: *Philosophy of Science*, Vol. 59, No. 1 (Mar., 1992), pp. 120-141
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/188122
Accessed: 18/01/2011 15:45

DIACHRONIC RATIONALITY*

PATRICK MAHER†‡

Department of Philosophy
University of Illinois at Urbana-Champaign

This is an essay in the Bayesian theory of how opinions should be revised over time. It begins with a discussion of the principle that van Fraassen has dubbed "Reflection". This principle is not a requirement of rationality; a diachronic Dutch argument, that purports to show the contrary, is fallacious. But under suitable conditions, it is irrational to actually implement shifts in probability that violate Reflection. Conditionalization and probability kinematics are special cases of the principle not to implement shifts that violate Reflection; hence these principles are also requirements of rationality under suitable conditions, though not universal requirements of rationality.

1. Reflection. Suppose you currently have a (personal) probability function p, and let R_q denote that at some future time $t + x$ you will have probability function q. Goldstein (1983) and van Fraassen (1984) have claimed that the following identity is a requirement of rationality (where "$\cdot$" stands for any proposition):[1]

$$p(\cdot|R_q) = q(\cdot).$$

Following van Fraassen (1984), I will refer to this identity as *Reflection*.

As an example of what Reflection requires, suppose you are sure that you cannot drive safely after having 10 drinks. Suppose further that you are sure that after 10 drinks, you would be sure (wrongly, as you now think) that you could drive safely. Then you violate Reflection. For if p is your current probability function, q the one you would have after 10 drinks, and D the proposition that you can drive safely after having 10 drinks, we have

$$p(D|R_q) \approx 0 < 1 \approx q(D).$$

Reflection requires $p(D|R_q) = q(D) \approx 1$. Thus you should now be sure

*Received June 1990; revised November 1990.

†This paper was written while I was a fellow in the Michigan Society of Fellows. The paper has benefited from comments by David Christensen, Howard Sobel, and Bas van Fraassen.

‡Send reprint requests to the author, Department of Philosophy, University of Illinois at Urbana-Champaign, 105 Gregory Hall, 810 South Wright St., Urbana, IL 61801, USA.

[1]Goldstein actually defends a stronger condition, but the argument for his stronger condition is the same as for the weaker one stated here.

Philosophy of Science, 59 (1992) pp. 120–141.

that you would not be in error if in the future you become sure that you can drive safely after having 10 drinks.

1.1. The Dutch Book Argument. Why should we think Reflection is a requirement of rationality? According to Goldstein and van Fraassen, this conclusion is established by a diachronic Dutch book argument. A diachronic Dutch book argument differs from a regular Dutch book argument in that the bets are not all offered at the same time. But like a regular Dutch book argument, it purports to show that anyone who violates the condition is willing to accept bets which together produce a sure loss and, hence, is irrational.

Since the diachronic Dutch book argument for Reflection has been stated in full generality elsewhere (Goldstein 1983, van Fraassen 1984, Skyrms 1987), I will here merely illustrate how it works. Suppose, then, that you violate Reflection with respect to drinking and driving in the way indicated above. For ease of computation, I will assume that $p(D) = 0$ and $q(D) = 1$. (Using less extreme values would not change the overall conclusion.) Let us further assume that your probability that you will have 10 drinks tonight is $^1/_2$. The Dutch bookie tries to make a sure profit from you by first offering a bet b_1 whose payoff in units of utility is

$$-2 \text{ if } DR_q; \quad 2 \text{ if } \bar{D}R_q; \quad -1 \text{ if } \bar{R}_q.$$

(Conjunction is represented by concatenation, and negation by overbars. For example, $\bar{D}R_q$ is the proposition that D is false and R_q is true.) For you at this time, $p(DR_q) = 0$, and $p(\bar{D}R_q) = p(\bar{R}_q) = {}^1/_2$. Thus the expected utility of b_1 is $^1/_2$. We are taking the utility of the status quo to be 0, and so the bookie figures that you will accept this bet. If you accept the bet, and do not get drunk (R_q is false), you lose 1 unit of utility. If you accept and do get drunk (R_q is true), the bookie offers you b_2, whose payoff in units of utility is

$$1 \text{ if } D; \quad -3 \text{ if } \bar{D}.$$

Since you are now certain D is true, accepting b_2 increases your expected utility, and so the bookie figures you will accept it. But now, if D is true, you gain 1 from b_2, but lose 2 from b_1, for an overall loss of 1. And if D is false, you gain 2 from b_1, but lose 3 from b_2, again losing 1 overall. Thus no matter what happens, you lose. (In presentations of this argument, it is usual to have two bets, where I have the single bet b_1. Those two bets would be a bet on R_q, and a bet on $\bar{D}$ which is called off if R_q is false. By using a single bet instead, I show that the argument does not here require the assumption that bets which are separately acceptable are also jointly acceptable.)

1.2. Counterexamples. Despite this argument, there are compelling prima facie counterexamples to Reflection. Indeed, the drinking/driving example is already a prima facie counterexample; it seems that you would be right to now discount any future opinions you might form while intoxicated—contrary to what Reflection requires. But we can make the counterexample even more compelling by supposing you are *sure* that tonight you will have 10 drinks. It then follows from Reflection that you should *now* be sure that you can drive safely after having 10 drinks:

$$\begin{aligned} \textit{Proof. } p(D) &= p(D|R_q), &&\text{since } p(R_q) = 1 \\ &= q(D), &&\text{by Reflection} \\ &= 1. \end{aligned}$$

This result seems plainly wrong. Nor does it help to say that a rational person should not drink so much, for it may be that the drinking you know you will do tonight will not be voluntary.

A defender of Reflection might try responding to such counterexamples by claiming that the person you would be when drunk is not the same person who is now sober. If you were absolutely sure of this, then for you $p(R_q) = 0$, since R_q asserts that *you* will come to have probability function q. (I assume that you are sure you cannot come to have q other than by drinking. This is a plausible assumption if, as we can suppose, q also assigns an extremely high probability to the proposition that you have been drinking.) In that case, $p(\cdot|R_q)$ may be undefined, and the counterexample thereby avoided. But this is a desperate move. Nobody I know gives any real credence to the claim that having 10 drinks, and as a result thinking they can drive safely, would destroy their personal identity. They are certainly not certain that this is true.

Alternatively, defenders of Reflection may bite the bullet and declare that even when it is anticipated that your probabilities will be influenced by drugs, Reflection should be satisfied. Perhaps nothing is too bizarre for such a die-hard defender of Reflection to accept. However, it may be worth pointing out a peculiar implication of the position here being embraced: It entails that rationality requires taking mind-altering drugs, in circumstances where that position seems plainly false. I will now show how that conclusion follows.

It is well known that under suitable conditions, gathering evidence increases the expected utility of subsequent choices, if it has any effect at all. The following conditions are sufficient:[2]

[2] I assume causal decision theory. For a discussion of this theory, and a proof that the stated conditions are indeed sufficient in this theory, see Maher (1990).

1. The evidence is "cost-free"; that is, gathering it does not alter what acts are subsequently available, nor is any penalty incurred merely by gathering the evidence.
2. Reflection is satisfied for the shifts in probability which could result from gathering the evidence. (Maher 1990 refers to Reflection as *Miller's principle*.) That is to say, if p is your current probability function, then for any probability function q you could come to have as a result of gathering the evidence, $p(\cdot|R_q) = q(\cdot)$.
3. The decision to gather the evidence is not "symptomatic"; that is, it is not probabilistically relevant to states it does not cause.
4. Probabilities satisfy the axioms of probability, and choices maximize expected utility at the time they are made.

Now suppose you have the opportunity of taking a drug which will influence your probabilities in some way which is not completely predictable. The drug is cost-free (in particular, it has no direct effect on your health or wealth), and the decision to take the drug is not symptomatic. Assume also that rationality requires condition 4 above to be satisfied. If Reflection is a general requirement of rationality, condition 2 should also be satisfied for the drug-induced shifts. Hence all four conditions are satisfied, and it follows that you cannot reduce your expected utility by taking this drug; and you may increase it.

For example, suppose a bookie is willing to bet with you on the outcome of a coin toss. You have the option of betting on heads or tails, and you receive \$1 if you are right, while losing \$2 if you are wrong. Currently your probability that the coin will land heads is $^1/_2$, and so you now think the best thing to do is not bet. (I assume that your utility function is roughly linear for such small amounts of money.) But suppose you can take a drug which will make you certain of what the coin toss will be; you do not know in advance whether it will make you sure of heads or tails, and you antecedently think both results equally likely. (This condition is necessary in order to ensure that your decision to take the drug is not symptomatic. If you thought the drug was likely to make you sure the coin will land heads, say, and if Reflection is satisfied, then the probability of the coin landing heads, given that you take the drug, would also be high. Since taking the drug has no causal influence on the outcome of the toss, and since the unconditional probability of heads is $^1/_2$, taking the drug would then be a symptomatic act.) The drug is cost-free, and you satisfy condition 4 above. Then if Reflection should hold with regard to the drug-induced shifts, you think you can make money by taking the drug. For after you take the drug, you will bet on the outcome you are then certain will result; and if you satisfy Reflection, you are now certain that bet will be successful. By contrast, if you do not take the

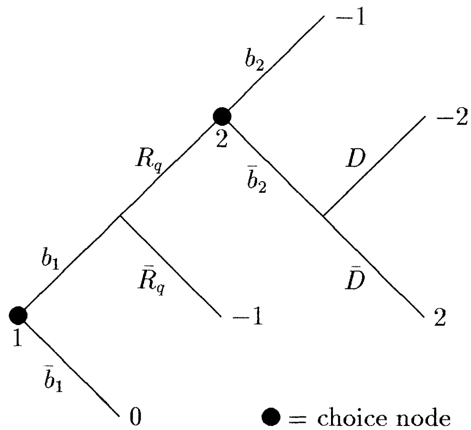

Figure 1. Decision tree of the Dutch book argument.

drug, you do not expect to make a profit betting on this coin toss. Thus the principle of maximizing expected utility requires you to take the drug.

But in fact, it is clear that taking the drug need not be rational. You could perfectly rationally think that the bet you would make after taking the drug has only a fifty-fifty chance of winning, and hence that taking the drug is equivalent to choosing randomly to bet on heads or tails. Since thinking that violates Reflection, we have another reason to deny that Reflection is a requirement of rationality.

1.3. The Fallacy. We now face a dilemma. On the one hand, we have a diachronic Dutch book argument to show that Reflection is a requirement of rationality. And on the other hand, we have strong reasons for saying that Reflection is *not* a requirement of rationality. There must be a mistake here somewhere. I argue that the mistake lies in the Dutch book argument for Reflection.

In the Dutch book argument for Reflection, the bets which together give you a sure loss are not offered at the same time. In the example of section 1.1, your decision on b_2 is made only after you have accepted b_1 and have shifted your probability function from p to q. Thus what you are faced with is a sequential decision problem. Figure 1 shows the decision tree for this example.

At node 1 in this tree, you think the rational choice to make at node 2, should you get there, would be to reject b_2. However, you also know that were you to get to node 2, you would then think it rational to accept b_2. Consequently, at node 1, you foresee that if you accept b_1 and R_q is true, you will lose 1 unit of utility; and you know that you will also lose

1 unit of utility if you accept b_1 and R_q is false. Thus the expected utility of choosing b_1 at node 1 is -1, while that of rejecting b_1 is zero. Hence as an expected utility maximizer, you are not actually willing to accept b_1.

There is no conflict here with the idea that your probabilities are defined by your preferences regarding bets. It may help to explain this with an analogy. Suppose you say your probability for rain tomorrow is $^1/_2$. I respond by offering you a bet in which you win 1 utile if it rains, and lose 1 if it does not. But suppose you know that come April 15, you will have to pay a tax on your winnings, while gambling losses are not tax deductible. In that case, it is consistent with the preference interpretation of probability for you to reject the bet, even though your probability for rain tomorrow is $^1/_2$, and I am offering you even odds on rain. The point is that while I am offering even odds, the bet is not at even odds for you once the tax situation is taken into account.

In the Dutch book argument for Reflection, being offered b_2 at node 2 is viewed by you at node 1 as like a tax, if you do not satisfy Reflection. Hence you can decline the bet b_1 for the same reason you would decline my bet on rain: When all things are considered, acceptance of these bets does not maximize expected utility. Bayesians, when careful, have always said that in making decisions, one needs to look ahead and take into account what may happen in the future. That is why Ramsey ([1926] 1978) took gambles to be between possible worlds. Similarly, Savage (1954, 15–17) enunciates a "Look before you leap" principle.

So we see that, contrary to the Dutch book argument, you can violate Reflection without being willing to accept a sure loss. Thus the Dutch book argument for Reflection is fallacious; it assumes that a rational person with probability function p must be willing to accept b_1, whereas a person who violates Reflection, and foresees the offer of b_2, will not accept b_1. This point has been made by Levi (1987, 204f.). I am here discussing what Levi calls "Case 2".

Let us consider some ways in which one might try to repair the argument. First, I have been assuming you are certain that, if you accept b_1 and R_q obtains, then b_2 will be offered. What if you are not certain of this? Since the expected utility of accepting b_1 is positive given that b_2 will not be offered, the expected utility of accepting b_1 also will be positive if the probability of b_2 being offered is sufficiently small. And if b_2 were in fact offered, you would then suffer a loss. However, now the loss is not a sure loss, and to risk a possible loss in order to obtain a reward is not irrational. (In the present case, it is also true that the bookie can, at will, make you suffer a loss. But this also does not convict you of irrationality. If I walk past my neighbor's house, he could shoot me, and thus I put myself in a position where he could at will make me suffer a loss; but going for a walk is not thereby irrational.) So this modification cannot repair the argument.

Another attempt to repair the argument would be to claim that rational persons do not anticipate making future choices which they now regard as not optimal choices in the relevant situation. Since the person who violates Reflection declines b_1 on the basis of such an anticipation, it would then follow that Reflection is a requirement of rationality. But this attempted repair will not work, for two reasons. First, the claim is false, since rational persons can anticipate that they might (perhaps involuntarily) take a mind-altering drug, and the choices which would be made under the influence of a drug need not now be regarded as optimal. Second, the claim is really just another way of stating the Reflection principles, so this attempted repair turns the Dutch book argument for Reflection into a circular argument.

I conclude that the Dutch book argument for Reflection has no cogency at all. Consequently, this argument provides no reason not to draw the obvious conclusion from the counterexamples in section 1.2: Reflection is not a requirement of rationality.[3]

1.4. Integrity. Recognizing the implausibility of saying Reflection is a requirement of rationality, van Fraassen (1984, 250–255) tried to bolster its plausibility with a voluntarist conception of personal probability judgements. He claimed that personal probability judgements express a kind of commitment; and he averred that integrity requires you to stand behind your commitments, including conditional ones. For example, he says your integrity would be undermined if you allowed that were you to promise to marry me, you still might not do it. And by analogy, he concludes that your integrity would be undermined if you said that your probability for A, given that tomorrow you give it probability r, is something other than r.

I agree that a personal probability judgement is a kind of commitment; to make such a judgement is to accept a constraint on your choices between uncertain prospects. For example, if you judge A to be more probable than B, and if you prefer \$1 to nothing, then faced with a choice between

(i) \$1 if A, nothing otherwise

and

[3]A different attack on the Dutch book argument for Reflection is made by Christensen (1991), who argues that it need not be irrational to be open to a diachronic Dutch book. I find his argument persuasive. However, the argument is not needed, because I have shown that violation of Reflection does not entail susceptibility to a diachronic Dutch book.

(ii) nothing if A, \$1 otherwise,

you are committed to choosing (i). But of course, you are not thereby committed to making this choice at all times in the future; you can revise your probabilities without violating your commitment. The commitment is to *now* make that choice, if *now* presented with those options. But this being so, a violation of Reflection is not analogous to thinking you might break a marriage vow. To think you might break a marriage vow is to think you might break a commitment. To violate Reflection is to not *now* be committed to acting in accord with a future commitment, on the assumption that you will in the future have that commitment. The difference is that in violating Reflection, you are not thereby conceding that you might ever act in a way that is contrary to your commitments at the time of action. A better analogy for violations of Reflection would be saying that you now think you would be making a foolish choice if you were to decide to marry me. In this case, as in the case of Reflection, you are not saying you could violate your commitments; you are merely saying that you do not now endorse those commitments, even on the supposition that you were to make them. Saying this does not undermine your status as a person of integrity.

1.5. Reflection and Learning. In the typical case of taking a mind-altering drug, Reflection is violated, and we also feel that while the drug would shift our probabilities, we would not have *learned* anything in the process. For instance, if a drug will make you certain of the outcome of a coin toss, then under typical conditions the shift produced by the drug does not satisfy Reflection, and one also does not regard taking the drug as a way of *learning* the outcome of the coin toss.

Conversely, in typical cases where Reflection is satisfied, we do feel that the shift in probabilities would involve learning something. For example, suppose Persi is about to toss a coin, and suppose you know that Persi can (and will) toss the coin so that it lands how he wants, and that he will tell you what the outcome will be if you ask. Then asking Persi about the coin toss will, like taking the mind-altering drug, make you certain of the outcome of the toss. But in this case, Reflection will be satisfied, and we can say that by asking Persi you will *learn* how the coin is going to land.

What makes the difference between these cases is not that a drug is involved in one, and testimony in the other. This can be seen by varying the examples. Suppose you think Persi really has no idea how the coin will land, but has such a golden tongue that if you talked to him you would come to believe him; in this case, a shift caused by talking to Persi will not satisfy Reflection, and you will not think that by talking to him you will learn the outcome of the coin toss (even though you will become sure of some outcome). Conversely, you might think that if you take the drug, a benevolent genie will influence the coin toss so that it agrees with what the drug would make you believe; in this case, the shift in probabilities caused by taking the drug will satisfy Reflection, and you will think that by taking the drug you will learn the outcome of the coin toss.

These considerations lead me to suggest that regarding a potential shift in probability as a learning experience is the same thing as satisfying Reflection in regard to that shift. Symbolically, you regard the shift from p to q as a learning experience just in case $p(\cdot|R_q) = q(\cdot)$.[4]

Shifts which do not satisfy Reflection, though not learning experiences in the sense just defined, may still involve some learning. For example, if q is the probability function you would have after taking the drug that makes you sure of the outcome of the coin toss, you may think that in shifting to q you would learn that you took the drug, but not learn the outcome of the coin toss. In general, what you think you would learn in shifting from p to q is represented by the difference between p and $p(\cdot|R_q)$. (This assumes that q records everything relevant about the shift, so that "what you learned in the shift from p to q" is a well-defined notion. The need for that assumption could be avoided by replacing R_q with a proposition that specifies your probability distribution at every instant between t and $t + x$.) When Reflection is satisfied, what is learned is represented by the difference between p and q, and we call the whole shift a learning experience.

Learning, so construed, is not limited to cases in which new empirical evidence is acquired. You may have no idea what is the square root of 289, but you may also think that if you pondered it long enough, you would come to concentrate your probability on some particular number, and that potential shift may well satisfy Reflection. In this case, you would regard the potential shift as a learning experience, though no new empirical evidence has been acquired. On the other hand, any shift in probability which is thought to be due solely to the influence of evidence is necessarily regarded as a learning experience. Thus satisfaction of Reflection is necessary, but not sufficient, for regarding a shift in probability as due to empirical evidence.

A defender of Reflection might think of responding to the counterexamples by limiting the principle to shifts of a certain kind. But the observations made in this section show that such a response will not help. If Reflection were said to be a requirement of rationality only for shifts caused in a certain way (e.g., by testimony rather than drugs), then there

[4]This proposal was suggested to me by Skyrms (1990), who assumes that what is thought to be a learning experience will satisfy Reflection. He calls Reflection, "Principle (M)".

would still be counterexamples to the principle. And if Reflection were said to be a requirement of rationality for shifts that are regarded as learning experiences, or as due to empirical evidence, then the principle would be one that it is impossible to violate, and hence vacuous as a principle of rationality.[5]

1.6. Reflection and Rationality. Although there is nothing irrational about violating Reflection, it is often irrational to actually implement those potential shifts which violate Reflection. That is to say, while one can rationally have $p(\cdot|R_q) \neq q(\cdot)$, it will in such cases often be irrational to choose a course of action that might result in acquiring the probability function q. The coin-tossing example of section 1.2 provides an illustration of this. Let H denote that the coin lands heads, and let q be the probability function you would have if you took the drug, and it made you certain of H. Then if you think taking the drug gives you only a random chance of making a successful bet, $p(H|R_q) = 0.5 < q(H) = 1$, and you violate Reflection; but then you would be irrational to take the drug since the expected return from doing so is $(^1/_2)(\$1) - (^1/_2)(\$2) < 0$.

This observation can be generalized, and made more precise, as follows. Let d and d' be two acts; for example, d might be the act of taking the drug in the coin-tossing case, and d' the act of not taking the drug. Assume that

(i) Any shift in probability after choosing d' would satisfy Reflection.

In the coin-tossing case, this will presumably be satisfied; if q' is the probability function you would have if you decided not to take the drug, q' will not differ much from p, and in particular $p(H|R_{q'}) = q'(H) = p(H) = 0.5$.

Assume also that

(ii) Acts d and d' influence expected utility only via their influence on what subsequent choices maximize expected utility.

More fully: Choosing d or d' may have an impact on your probability function, and thereby influence your subsequent choices; but (ii) requires that they not influence expected utility in any other way. So there must not be a reward or penalty attached directly to having any of the prob-

[5]Jeffrey (1988, 233) proposed to restrict Reflection to shifts that are "reasonable", without saying what that means. His proposal faces precisely the dilemma I have just outlined. If a "reasonable" shift is defined by its causal origin, Jeffrey's principle is not a requirement of rationality. If a "reasonable" shift is defined to be a learning experience, Jeffrey's principle is vacuous. In the next section, we will see that if a "reasonable" shift is a shift that it would be rational to implement, Jeffrey's principle is again not a requirement of rationality.

ability functions which could result from choosing d or d'; nor can the choice of d or d' alter what subsequent options are available. This condition will also hold in the coin-tossing example if the drug is free and has no deleterious effects on health, and otherwise if the situation is fairly normal.

Assume further that

(iii) If anything would be learned about the states by choosing d, it would also be learned by choosing d'.

What I mean by (iii) is that the following four conditions are all satisfied. Here Q is the set of all probability functions which you could come to have if you chose d:

(a) You are sure there is a fact about what probability function you would have if you chose d, that is, you give probability 1 to the proposition that for some q, the counterfactual conditional $d \rightarrow R_q$ is true.
(b) For all $q \in Q$ there is a probability function q' such that $p(d' \rightarrow R_{q'}|d \rightarrow R_q) = 1$.
(c) There is a set S of states of nature which are suitable for calculating the expected utility of the acts which will be available after the choice between d and d' is made. (What this requires is explained in the first paragraph of the Appendix.)
(d) For all $q \in Q$, and for q' related to q as in (b), and for all $s \in S$, $p(s|R_q) = p(s|R_{q'})$.

In the coin-tossing example, condition (a) can be assumed to hold: Presumably the drug is deterministic so that there is a fact about what probability function you would have if you took the drug, though you do not know in advance what that fact is. Condition (b) holds trivially in the coin-tossing example because not taking the drug would leave you with the same probability function q', regardless of what effect the drug would have. Condition (c) is satisfied by taking $S = \{H, \bar{H}\}$. And it is a trivial exercise to show that (d) holds, since

$$p(H|R_q) = p(H) = 1/2 = p(H|R_{q'}).$$

The coin-tossing example thus satisfies condition (iii). We could say that in this example, you learn nothing about the states, whether you choose d or d'.

Also assume that

(iv) Acts d and d' have no causal influence on the states S mentioned in (c).

In the coin-tossing example, neither taking the drug, nor refusing it, has any causal influence on how the coin lands; and so (iv) is satisfied.

Finally, assume that

(v) Acts d and d' are not evidence for events they have no tendency to cause.

In the coin-tossing example, (iv) and (v) together entail that $p(H|d) = p(H|d') = 1/2$, which is what one would expect to have in this situation.

THEOREM. *If conditions (i)–(v) are known to hold, then the expected utility of d' is not less than that of d, and may be greater.*

So it would always be rational to choose d', but it may be irrational to choose d. The proof is given in the Appendix.

The theorem can fail when the stated conditions do not hold. For one example of this, suppose you are convinced that a superior being gives eternal bliss to all and only those who are certain that pigs can fly. Suppose also that a particular drug, if you take it, will make you certain that pigs can fly. If q is the probability function you would have after taking this drug, and F is the proposition that pigs can fly, then $q(F) = 1$. Presumably $p(F|R_q) = p(F) \approx 0$. So the shift resulting from taking this drug violates Reflection. On the other hand, not taking the drug would leave your current probability essentially unchanged. But in view of the reward attached to being certain pigs can fly, it would (or at least, could) be rational to take the drug, and thus implement a violation of Reflection. (If eternal bliss includes epistemic bliss, taking the drug could even be rational from a purely epistemic point of view.) Here the result fails, because condition (ii) does not hold: Taking the drug influences your utility other than via its influence on your subsequent decisions.

To illustrate another way in which the result may fail, suppose you now think there is a 90 percent chance that Persi knows how the coin will land, but that after talking to him you would be certain that what he told you was true. Again letting H denote that the coin lands heads, and letting q_H be the probability function you would have if Persi told you the coin will land heads, we have $p(H|R_{q_H}) = 0.9$, while $q_H(H) = 1$. Similarly for $q_{\bar{H}}$. Thus talking to Persi implements a shift which violates Reflection. If you do not talk to Persi, you will have probability function q' which, so far as H is concerned, is identical to your current probability function p; so $p(H|R_{q'}) = q'(H) = 0.5$. Thus not talking to Persi avoids implementing a shift which violates Reflection. Your expected return from talking to Persi is

$$(0.9)(\$1) + (0.1)(-\$2) = \$0.70.$$

Since you will not bet if you do not talk to Persi, the expected return from not talking to him is zero. Hence talking to Persi maximizes your expected monetary return. And assuming your utility function is approximately linear for small amounts of money, it follows that talking to Persi maximizes expected utility. Here the theorem fails because condition (iii) fails. By talking to Persi, you do learn something about how the coin will land; and you learn nothing about this if you do not talk to him. The theorem I stated implies that the expected utility of talking to Persi is no higher than that of learning what you would learn from him, without violating Reflection; but in the problem I have described, the latter option is not available.

I will summarize the above theorem by saying that, *other things being equal,* implementing a shift which violates Reflection cannot have greater expected utility than implementing a shift which satisfies Reflection. Conditions (i)–(v) specify what is meant here by "other things being equal". This, not the claim that a rational person must satisfy Reflection, gives the true connection between Reflection and rationality.

2. Conditionalization. Bayesian theory incorporates a theory of learning, and presentations of this theory of learning typically give a central place to a principle of *conditionalization.* We can state the principle as follows:

> Conditionalization: If your current probability function is p, and if q is the probability function you would have if you learned E and nothing else, then $q(\cdot)$ should be identical to $p(\cdot|E)$.

An alternative formulation, couched in terms of evidence rather than learning, will be discussed in section 2.3.

Paul Teller (1973, 1976) reports a Dutch book argument due to David Lewis, which purports to show that conditionalization is a requirement of rationality. The argument is essentially the same as the Dutch book argument for Reflection (but this way of putting the matter reverses the chronological order, since Lewis formulated the argument for conditionalization before the argument for Reflection was advanced), and is fallacious for the same reason.

2.1. Conditionalization, Reflection, and Rationality. In this section, I will argue that conditionalization is not a universal requirement of rationality, and will explain what I take to be its true normative status.

Recall what the conditionalization principle says: If you learn E, and nothing else, then your posterior probability should equal your prior probability conditioned on E. But what does it mean to "learn E, and nothing else"? In section 1.5, I suggested that what you think you would learn in shifting from p to q is represented by the difference between p and

$p(\cdot|R_q)$. From this perspective, we can say that you think you would learn E and nothing else, in shifting from p to q, just in case $p(\cdot|R_q) = p(\cdot|E)$.

This is only a subjective account of learning; it gives an interpretation of what it means to *think E* would be learned, not what it means to *really* learn E. But I have no idea how to make sense of the latter notion, except as a projection of the former. And if we could make sense of a truly objective notion of learning, then presumably it would be possible to learn E without believing you had learned it; in which case, there is no plausibility in claiming that you rationally ought to conditionalize on E. (I take it that you are irrational if you violate your own standards, and there need be no such violation in this case.) Consequently, we maximize both clarity and charity if we take the "learning" referred to in the principle of conditionalization to be learning as judged by you. In what follows, I use the term "learning" in this way.

So if you learned E and nothing else, and if your probabilities shifted from p to q, then $p(\cdot|R_q) = p(\cdot|E)$. If you also satisfy Reflection in regard to this shift, then $p(\cdot|R_q) = q(\cdot)$, and so $q(\cdot) = p(\cdot|E)$, as conditionalization requires. This simple inference shows that Reflection entails conditionalization.

It is also easy to see that if you learn E, and nothing else, and if your probabilities shift in a way that violates Reflection, then your probability distribution is not updated by conditioning on E. For since you learned E, and nothing else, $p(\cdot|R_q) = p(\cdot|E)$; and since Reflection is not satisfied in this shift, $q(\cdot) \neq p(\cdot|R_q)$, whence $q(\cdot) \neq p(\cdot|E)$.

These results together show that conditionalization is equivalent to the following principle: When you learn E and nothing else, do not implement a shift which violates Reflection. But we saw, in section 1.6, that in some cases it is rational to implement a shift which violates Reflection. I will now show that some of these cases are ones in which you learn E, and nothing else. This suffices to show that it can be rational to violate conditionalization.

Consider again the situation in which you are sure there is a superior being who will give you eternal bliss, if and only if you are certain that pigs can fly; and there is a drug available which will make you certain of this. Let d be the act of taking the drug, and q the probability function you would have after taking the drug. Then we can plausibly suppose that $p(\cdot|R_q) = p(\cdot|d)$, and hence that in taking the drug you learn d, and nothing else. Consequently, conditionalization requires that your probability function after taking the drug be $p(\cdot|d)$, which it will not be. (With F denoting that pigs can fly, $p(F|d) = p(F) \approx 0$, while $q(F) = 1$.) Hence taking the drug implements a violation of conditionalization. Nevertheless, it is rational to take the drug in this case, and hence to violate conditionalization.

Similarly for the other example of section 1.6. Here you think there is a 90 percent chance that Persi knows how the coin will land, but you know that after talking to him, you would become certain that what he told you was true. We can suppose that in talking to Persi, you think you will learn what he said, and nothing else. Then an analysis just like that given for the preceding example shows that talking to Persi implements a violation of conditionalization. Nevertheless, it is rational to talk to Persi, because (as we saw) this maximizes your expected utility.

It is true that in both of these examples, there are what we might call "extraneous" factors that are responsible for the rationality of violating conditionalization. In the first example, the violation is the only available way to attain eternal bliss; and in the second example, it is the only way to acquire some useful information. Can we show that, putting aside such considerations, it is irrational to violate conditionalization? Yes, we have already proven that. For we saw that when other things are equal (in a sense made precise in section 1.6), expected utility can always be maximized without implementing a violation of Reflection. As an immediate corollary, we have that, when other things are equal, expected utility can always be maximized without violating conditionalization.[6]

To summarize, the principle of conditionalization is a special case of the principle which says not to implement shifts that violate Reflection. Like that more general principle, it is not a universal requirement of rationality, but it is a rationally acceptable principle in contexts where other things are equal in the sense made precise in section 1.6.

2.2. *Other Arguments for Conditionalization.* Lewis's Dutch book argument is not the only argument which has been advanced to show that conditionalization is a requirement of rationality. What I have said in the preceding section implies that these other arguments must also be incorrect. I will show that this is so for arguments offered by Teller and by Howson.

After presenting Lewis's Dutch book argument, Teller (1973, 1976) proceeds to offer an argument of his own for conditionalization. The central assumption of this argument is that if you learn E and nothing else, then for all propositions A and B which entail E, if $p(A) = p(B)$, then it ought to be the case that $q(A) = q(B)$. (Here, as before, p and q are your

[6]Brown (1976) gives a direct proof of a less general version of this result. What makes his result less general is that it only applies to cases where for each E you might learn, there is a probability function q such that you are sure q would be your probability function if you learned E, and nothing else. This means that Brown's result is not applicable to the coin-tossing example of section 1.2, for example. (In this example, your posterior probability, on learning that you took the drug, could give probability 1 to either heads or tails.) Another difference between Brown's proof and mine is that his does not apply to probability kinematics (see section 3).

prior and posterior probability functions, respectively.) Given this assumption, Teller is able to derive the principle of conditionalization. But the counterexamples which I have given to conditionalization are also counterexamples to Teller's assumption. To see this, consider the first counterexample in which taking a drug will make you certain that pigs can fly, and this will give you eternal bliss. Let F and d be as before, and let G denote that the moon is made of green cheese. We can suppose that in this example, $p(Fd) = p(Gd)$, and $q(Fd) = q(F) > q(G) = q(Gd)$. Assuming that d is all you learn from taking the drug, we have a violation of Teller's principle. But the shift from p to q involves no failure of rationality. You do not want $q(F)$ to stay small, or else you will forgo eternal bliss; nor is there any reason to become certain of G, and preserve Teller's principle that way. Thus Teller's principle is not a universal requirement of rationality, and hence his argument fails to show that conditionalization is such a requirement. (My second counterexample to conditionalization could be used to give a parallel argument for this conclusion.)

Perhaps Teller did not intend his principle to apply to the sorts of cases considered in my counterexamples. If so, there may be no dispute between us since I have agreed that conditionalization is rational when other things are equal. But then I would say that Teller's defense of conditionalization is incomplete because he gives no method for distinguishing the circumstances in which his principle applies. By contrast, the decision-theoretic approach I have used makes it a straightforward matter of calculation to determine under what circumstances rationality requires conditionalization.

I turn now to Howson's argument for conditionalization. Howson interprets $p(H)$ as the betting quotient on H which you now regard as fair, $p(H|E)$ as the betting quotient which you now think would be fair were you to learn E (and nothing else), and $q(H)$ as the betting quotient which you will in fact regard as fair after learning E (and nothing else). His argument is the following (with my notation):

> $p(H|E)$ is, as far as you are concerned, just what the fair betting quotient would be on H were E to be accepted as true. Hence from the knowledge that E is true you should infer (and it is an inference endorsed by the standard analyses of subjunctive conditionals) that the fair betting quotient on H is equal to $p(H|E)$. But the fair betting quotient on H after E is known is by definition $q(H)$. (Howson and Urbach 1989, 68)

I would not endorse Howson's conception of conditional probability. However, even granting Howson this conception, his argument is fallacious. Howson's argument rests on an assumption of the following form: People who accept "If A then B" are obliged by logic to accept B if they learn A. But this is a mistake; on learning A you might well decide to abandon the conditional "If A then B", thereby preserving logical consistency in a different way.

In the case at hand, Howson's conception of conditional probability says that you accept the conditional "If I were to learn E and nothing else, then the fair betting quotient for H would be $p(H|E)$". Howson wants to conclude from this that if you do learn E and nothing else, then logic obliges you to accept that the fair betting quotient for H is $p(H|E)$. But as we have seen, this does not follow; for you may reject the conditional. In fact, if you adopt a posterior probability function q, then your conditional probability for H becomes $q(H|E) = q(H)$; and according to Howson, this means you now accept the conditional "If I were to learn E and nothing else, then the fair betting quotient for H would be $q(H)$". In cases where conditionalization is violated, $q(H) \neq p(H|E)$, and so the conditional you now accept is different from the one you accepted before learning E.

Thus neither Teller's argument nor Howson's refutes my claim that it is sometimes rational to violate conditionalization. And neither is a substitute for my argument that, when other things are equal, rationality never requires violating conditionalization.

2.3. Van Fraassen on Conditionalization. In a recent article, van Fraassen (forthcoming) has argued that conditionalization is not a requirement of rationality. From the perspective of this paper, that looks at first sight to be a paradoxical position for him to take. I have argued that conditionalization is a special case of the principle not to implement shifts that violate Reflection. If this is accepted, then van Fraassen's claim that Reflection is a requirement of rationality implies that conditionalization is also a requirement of rationality.

I think the contradiction here is merely apparent. Van Fraassen's idea of how you could rationally violate conditionalization is that you might think that when you get some evidence and deliberate about it, you could have some unpredictable insight which will have the result that your posterior probability will differ from your prior conditioned on the evidence. Now I would say that if you satisfy Reflection, your unpredictable insight will be part of what you learned from this experience, and there is no violation of conditionalization. But there is a violation of what we could call

> Evidence-conditionalization: If your current probability function is p, and if q is the probability function you would have if you acquired evidence E and no other evidence, then $q(\cdot)$ should be identical to $p(\cdot|E)$.

This principle differs from conditionalization as I defined it, in having E be the total *evidence* acquired, rather than the totality of what was *learned.* These are different things because, as argued in section 1.5, not all learning involves getting evidence. Where ambiguity might otherwise arise, we could call conditionalization as I defined it *learning-conditionalization.*

These two senses of conditionalization are not usually distinguished in discussions of Bayesian learning theory, presumably because those discussions tend to focus on situations in which it is assumed that the only learning that will occur is due to acquisition of evidence. But once we consider the possibility of learning without acquisition of evidence, evidence-conditionalization becomes a very implausible principle. For example, suppose you were to think about the value of $\sqrt{289}$, and that as a result you substantially increase your probability that it is 17. We can suppose that you acquired no evidence over this time, in which case evidence-conditionalization would require your probability function to remain unchanged. Hence if evidence-conditionalization were a correct principle, you would have been irrational to engage in this ratiocination. This is a plainly false conclusion. (On the other hand, there need be no violation of learning-conditionalization; you may think you *learned* that $\sqrt{289}$ is 17.)

So van Fraassen is right to reject evidence-conditionalization, and doing so is not inconsistent with his endorsement of Reflection. But that endorsement of Reflection does commit him to learning-conditionalization, and I have urged that this principle should also be rejected.

3. Probability Kinematics. It is possible for the shift from p to q to satisfy Reflection, without it being the case that there is a proposition E such that $q(\cdot) = p(\cdot|E)$. When this happens, you think you have learned something, but there is no proposition E which expresses what you learned. The principle of conditionalization is then not applicable.

Jeffrey (1983, chap. 11) proposed a generalization of conditionalization, called probability kinematics, that applies in such cases. Jeffrey supposed that what was learned can be represented as a shift in the probability of the elements of some partition $\{E_i\}$. The rule of probability kinematics then specifies that the posterior probability function q be related to the prior probability p by the condition

$$q(\cdot) = \sum_i p(\cdot|E_i)q(E_i).$$

Armendt (1980) has given a Dutch book argument to show that the rule of probability kinematics is a requirement of rationality. This argument, however, has the same fallacy as the Dutch book arguments for Reflection and conditionalization. Furthermore, my account of the true status of conditionalization also extends immediately to probability kinematics.

A natural interpretation of what it means for you to think what you learned is represented by a shift from p to q' on the E_i would be that the shift is to q, and

$$p(\cdot|R_q) = \sum_i p(\cdot|E_i)q'(E_i).$$

But then it follows that the requirement to update your beliefs by probability kinematics is equivalent to the requirement not to implement any shifts which violate Reflection. Hence updating by probability kinematics is not in general a requirement of rationality, though it is a rational principle when other things are equal, in the sense of section 1.6.

4. Conclusion. If diachronic Dutch book arguments were sound, then Reflection, conditionalization, and probability kinematics would all be requirements of rationality. But these arguments are fallacious, and, in fact, none of these three principles is a general requirement of rationality. Nevertheless, there is some truth to the idea that these three principles are requirements of rationality. Bayesian decision theory entails that when other things are equal, rationality never requires implementing a shift in probability that violates Reflection. Conditionalization and probability kinematics are special cases of the principle not to implement shifts that violate Reflection. Hence we also have that when other things are equal, it is always rationally permissible, and may be obligatory, to conform to conditionalization and probability kinematics.

APPENDIX

This Appendix proves the theorem stated in section 1.6.

Given a suitable set X of states of nature (I will assume that the set of states is at most countable; this assumption can be removed by replacing summation with integration), the expected utility of any act a can be written as follows:

$$EU(a) = \sum_{x \in X} p(x)u(xa)$$

where p and u are the person's probability and utility functions, and xa denotes the conjunction of x and a. According to causal decision theory, which I assume here (for a discussion of this theory and its alternative, see, for example, Maher 1987), X is a suitable set of states for calculating the expected utility of a iff X is a partition such that for each $x \in X$, x is not causally influenced by a, and x determines the consequence which will be obtained if a is chosen. Consequences are here understood as including every aspect of the outcome that matters to the person.

Let S be a suitable set of states of nature for calculating the expected utility of the acts which will be available after the choice between d and d' is made. By (iv) the states $s \in$

S are causally independent of d, and hence conjunctions of the form $s.d \rightarrow R_q$ are also causally independent of d. (Where ambiguity would otherwise arise, I use a dot to represent conjunction; the scope extends to the end of the formula or to the next dot, if there is one. So $s.d \rightarrow R_q$ denotes the conjunction of s and $d \rightarrow R_q$.) Let the set of acts available after choosing d or d' be B, and assume there is a unique $b_q \in B$ which you would choose if your probability function were q. Then d together with $d \rightarrow R_q$ determines that you will choose b_q, and this together with s determines the unique consequence you will obtain as a result. Hence conjunctions of the form $s.d \rightarrow R_q$ are both causally independent of d and determine the consequence which will be eventually obtained as an indirect result of choosing d. We can therefore take these propositions to be our states for the purpose of computing the expected utility of d. Letting Q be the set of all probability functions which you could have after choosing d, we then have

$$EU(d) = \sum_{s \in S} \sum_{q \in Q} p(s.d \rightarrow R_q)u(sb_q). \tag{1}$$

You know that if d and $d \rightarrow R_q$ are true, then R_q must obtain. Hence $p(R_q|s.d \rightarrow R_q.d) = 1$. Thus (1) implies

$$EU(d) = \sum_{s \in S} \sum_{q \in Q} p(s.d \rightarrow R_q)p(R_q|s.d \rightarrow R_q.d)u(sb_q).$$

Applying Bayes's theorem to $p(R_q|s.d \rightarrow R_q.d)$ then gives

$$EU(d) = \sum_{s \in S} \sum_{q \in Q} p(s.d \rightarrow R_q)[p(s.d \rightarrow R_q|R_q d)p(R_q|d)]/[p(s.d \rightarrow R_q|d)]u(sb_q).$$

By (v), $p(s.d \rightarrow R_q|d) = p(s.d \rightarrow R_q)$, and so we have

$$EU(d) = \sum_{s \in S} \sum_{q \in Q} p(s.d \rightarrow R_q|R_q d)p(R_q|d)u(sb_q). \tag{2}$$

By condition (a), you are sure one of the $d \rightarrow R_q$ holds, and so $p(s.d \rightarrow R_q|R_q d) = p(s|R_q d)$. I assume that $p(d|R_q) = 1$, for all $q \in Q$; a sufficient condition for this would be that you know what act you have chosen after you choose it. Hence $p(s|R_q d) = p(s|R_q)$, and (2) simplifies to

$$EU(d) = \sum_{s \in S} \sum_{q \in Q} p(s|R_q)p(R_q|d)u(sb_q)$$

$$= \sum_{q \in Q} p(R_q|d) \sum_{s \in S} p(s|R_q)u(sb_q).$$

Since you are sure one of the $d \rightarrow R_q$ is true, $p(R_q|d) = p(d \rightarrow R_q|d)$; and by (v), $p(d \rightarrow R_q|d) = p(d \rightarrow R_q)$. Thus

$$EU(d) = \sum_{q \in Q} p(d \rightarrow R_q) \sum_{s \in S} p(s|R_q)u(sb_q). \tag{3}$$

Condition (b) asserts that for each $q \in Q$ there is a probability function q' such that $p(d' \rightarrow R_{q'}|d \rightarrow R_q) = 1$. Obviously, only one q' can satisfy this condition for a given q; I will use the notation $\phi(q)$ to denote this q'. I will also use Q' to denote the set of probability functions you might have after choosing d'. Then we can rearrange the summation over Q in (3), to give

$$EU(d) = \sum_{q' \in Q'} \sum_{\phi(q)=q'} p(d \rightarrow R_q) \sum_{s \in S} p(s|R_q)u(sb_q).$$

By condition (d), $p(s|R_q) = p(s|R_{\phi(q)})$. Also, condition (i) entails that $p(s|R_{\phi(q)}) = \phi(q)(s)$. Hence

$$\begin{aligned} EU(d) &= \sum_{q' \in Q'} \sum_{\phi(q)=q'} p(d \rightarrow R_q) \sum_{s \in S} \phi(q)(s)u(sb_q) \\ &\leq \sum_{q' \in Q'} \sum_{\phi(q)=q'} p(d \rightarrow R_q) \max_{b \in B} \sum_{s \in S} \phi(q)(s)u(sb) \\ &= \sum_{q' \in Q'} \left[\sum_{\phi(q)=q'} p(d \rightarrow R_q) \right] \max_{b \in B} \sum_{s \in S} q'(s)u(sb). \end{aligned} \tag{4}$$

By the definition of ϕ, $p(d' \rightarrow R_{q'}|d \rightarrow R_q)$ equals 1 if $q' = \phi(q)$, and 0 otherwise. Hence

$$\sum_{\phi(q)=q'} p(d \rightarrow R_q) = \sum_{q \in Q} p(d' \rightarrow R_{q'}|d \rightarrow R_q)p(d \rightarrow R_q).$$

The theorem of total probability then gives

$$\sum_{\phi(q)=q'} p(d \rightarrow R_q) = p(d' \rightarrow R_{q'}).$$

Substituting in (4), we have

$$EU(d) \leq \sum_{q' \in Q'} p(d' \rightarrow R_{q'}) \max_{b \in B} \sum_{s \in S} q'(s)u(sb). \tag{5}$$

The reasoning leading to (3) can be repeated, *mutatis mutandis*, for d', giving

$$EU(d') = \sum_{q' \in Q'} p(d' \rightarrow R_{q'}) \sum_{s \in S} p(s|R_{q'})u(sb_{q'}).$$

From condition (i) we have $p(s|R_{q'}) = q'(s)$, and so

$$\begin{aligned} EU(d') &= \sum_{q' \in Q'} p(d' \rightarrow R_{q'}) \sum_{s \in S} q'(s)u(sb_{q'}) \\ &= \sum_{q' \in Q'} p(d' \rightarrow R_{q'}) \max_{b \in B} \sum_{s \in S} q'(s)u(sb). \end{aligned} \tag{6}$$

Comparing (5) and (6), we see that $EU(d) \leq EU(d')$.

REFERENCES

Armendt, B. (1980), "Is There a Dutch Book Argument for Probability Kinematics?", *Philosophy of Science 47*: 583–588.

Brown, P. M. (1976), "Discussion: Conditionalization and Expected Utility", *Philosophy of Science 43*: 415–419.

Christensen, D. (1991), "Clever Bookies and Coherent Beliefs", *Philosophical Review 50*: 229–247.

Goldstein, M. (1983), "The Prevision of a Prevision", *Journal of the American Statistical Association 78*: 817–819.

Howson, C. and Urbach, P. (1989), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Jeffrey, R. C. (1983), *The Logic of Decision*. 2d ed. Chicago: University of Chicago Press.

———. (1988), "Conditioning, Kinematics, and Exchangeability", in B. Skyrms and W. L. Harper (eds.), *Causation, Chance, and Credence*, vol. 1. Dordrecht: Kluwer, pp. 221–255.

Levi, I. (1987), "The Demons of Decision", *The Monist 70*: 193–211.

Maher, P. (1987), "Causality in the Logic of Decision", *Theory and Decision 22*: 155–172.

———. (1990), "Symptomatic Acts and the Value of Evidence in Causal Decision Theory", *Philosophy of Science 57*: 479–498.

Ramsey, F. P. ([1926]1978), "Truth and Probability", republished in *Foundations*. Edited by D. H. Mellor. Atlantic Highlands, NJ: Humanities Press, pp. 58–100.

Savage, L. J. (1954), *The Foundations of Statistics*. New York: Wiley.

Skyrms, B. (1987), "Dynamic Coherence and Probability Kinematics", *Philosophy of Science 54*: 1–20.

———. (1990), "The Value of Knowledge", in C. W. Savage (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 14, *Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 245–266.

Teller, P. (1973), "Conditionalization and Observation", *Synthese 26*: 218–258.

———. (1976), "Conditionalization, Observation, and Change of Preference", in W. Harper and C. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. 1. Dordrecht: Reidel, pp. 205–253.

van Fraassen, B. C. (1984), "Belief and the Will", *Journal of Philosophy 81*: 235–256.

———. (forthcoming), "Rationality Does Not Require Conditionalization", in E. Ullmann-Margalit (ed.), *The Israel Colloquium Studies in the History, Philosophy, and Sociology of Science*, vol. 5. Dordrecht: Kluwer.