5

Diachronic rationality

We have seen that Bayesian confirmation theory rests on two assumptions: (1) That rational scientists have probabilities for scientific hypotheses, and (2) the principle of conditionalization. The latter is a *diachronic* principle of rationality, because it concerns how probabilities at one time should be related to probabilities at a later time.

Chapters 1–4 gave an extended argument in support of (1). This chapter will examine what can be said on behalf of (2). I will reject the common Bayesian view that conditionalization is a universal requirement of rationality, but argue that nevertheless it should hold in normal scientific contexts.

I begin by discussing a putative principle of rationality known as Reflection. A correct understanding of the status of this principle will be the key to my account of the status of conditionalization.

5.1 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:[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 ten drinks. Suppose further that you are sure that after ten drinks, you would

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

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 ten drinks, and D the proposition that you can drive safely after having ten 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 that you would not be in error, if in the future you become sure that you can drive safely after having ten drinks.

5.1.1 *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 that 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 1987b), 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 ten 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]

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

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

[2]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.

will accept this bet. If you accept the bet and do not get drunk (R_q is false), you lose one 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

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

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.[3]

5.1.2 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 ten drinks. It then follows from Reflection that you should *now* be sure that you can drive safely after having ten drinks.

Proof.

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

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

[3] 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 that are separately acceptable are also jointly acceptable.

absolutely sure of this, then for you $p(R_q) = 0$, since R_q asserts that *you* will come to have probability function q.[4] 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 ten drinks, and as a result thinking he or she can drive safely, would destroy his or her personal identity. They are certainly not sure that this is true.

Alternatively, defenders of Reflection may bite the bullet, and declare that even when it is anticipated that one's 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. But 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:[5]

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 that could result from gathering the evidence. 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.

[4] 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.

[5] 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 1990b). In that work, I referred to Reflection as *Miller's principle*.

Now suppose you have the opportunity of taking a drug that will influence your probabilities in some way that 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; you receive $1 if you are right and lose $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 that 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.[6] The drug is cost free, and you satisfy condition 4. 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 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 50-50 chance of winning, and hence that taking the drug is equivalent to choosing

[6] 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.

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.

5.1.3 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.

In (Maher 1992), following Levi (1987), I argued that a sophisticated bettor who looks ahead will not accept the bets offered in the Dutch book argument for Reflection. The thought was that if you look ahead, you will see that accepting b_1 inevitably leads to a sure loss, and hence will refuse to take the first step down the primrose path. This diagnosis assumed that if you do not accept b_1, you will not be offered b_2. However, Skyrms (in press) points out that the bookie could offer b_2 if R_q obtains, regardless of whether b_1 has been accepted. Faced with this strategy, you do best (maximize expected utility) to accept b_1 as well, and thus ensure a sure loss.

So with Skyrms's emendation, the diachronic Dutch book argument does show that if you violate Reflection, you can be made to suffer a sure loss. Yet as Skyrms himself agrees, it is not necessarily rational to conform to Reflection. Thus we have to say that *susceptibility to a sure loss does not prove irrationality.* This conclusion may appear counterintuitive; but that appearance is an illusion, I will now argue.

We say that act a *dominates* act b if, in every state, the consequence of a is better than that of b. It is uncontroversial that it is irrational to choose an act that is dominated by some other available act. Call such an act *dominated.* One might naturally suppose that accepting a sure loss is a dominated act, and thereby irrational.

But consider this case: I have bet that it will not rain today. The deal, let us say, is that I lose \$1 if it rains and win \$1 otherwise. How I came to make this bet does not matter – perhaps it looked attractive to me at the time; perhaps I made

	No rain	Rain
Accept 2nd bet	−\$0.50	−\$0.50
Don't accept	\$1	−\$1

Figure 5.1: Available options after accepting bet against rain

it under duress. In any case, storm clouds are now gathering and I think I will lose the bet. I would now gladly accept a bet that pays me \$0.50 if it rains and in which I pay \$1.50 otherwise. If I did accept such a new bet, then together with the one I already have, I would be certain to lose \$0.50. So I am willing to accept a sure loss. But I am not thereby irrational. The sure loss of \$0.50 is better than a high probability of losing \$1. Note also that although I am willing to accept a sure loss, I am *not* willing to accept a dominated option. My options are shown in Figure 5.1. The first option, which gives the sure loss, is not dominated by the only other available act.

So we see that acceptance of a sure loss is not always a dominated act; and when it is not, acceptance of a sure loss can be rational. I suggest that the intuitive irrationality of accepting (or being willing to accept) a sure loss results from the false supposition that acceptance of a sure loss is always a dominated option, combined with the correct principle that it is irrational to accept (or be willing to accept) a dominated option.

Let us apply this to the Dutch book argument for Reflection. In the example of Section 5.1.1, you are now certain that accepting b_2 would result in a loss, and hence you prefer that you not accept it. However, you also know that you will accept it if you get drunk. This indicates that your willingness to accept b_2 when drunk is not something you are now able to reverse (for if you could, you would). Thus you are in effect now stuck with the fact that you will accept b_2 if you get drunk, that is, if R_q is true. Hence you are in effect now saddled with the bet

$$1 \text{ if } DR_q; \qquad -3 \text{ if } \bar{D}R_q,$$

	DR_q	$\bar{D}R_q$	$\bar{R}_q$
Accept b_1	−1	−1	−1
Reject b_1	1	−3	0

Figure 5.2: Available options for Reflection violator

though it looks unattractive to you now. (This is analogous to the first bet in the rain example.) But you do have a choice about whether or not to accept b_1. Since b_1 pays

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

your options and their payoffs are as in Figure 5.2. Accepting b_1 ensures that you suffer a sure loss; but it is not a dominated option. In fact, since $p(D) = 0$ and $p(R_q) = 1/2$, accepting b_1 reduced your expected loss from -1.5 to -1. So in this case, as in the rain example, the willingness to accept a sure loss does not involve willingness to accept a dominated option, and does not imply irrationality.

If there is any irrationality in this case, it lies in the potential future acceptance of b_2. But because that future acceptance is outside your present control, it is no reason to say that you are now irrational. Perhaps your future self would be irrational when drunk, but that is not our concern. Reflection is a condition on your present probabilities only, and what we have seen is that you are not irrational to now have the probabilities you do, even though having these probabilities means you are willing to accept a sure loss.

Let us say that a Dutch book *theorem* asserts that violation of some condition leaves one susceptible to a sure loss, while a Dutch book *argument* infers from the theorem that violation of the condition is irrational. In Section 4.6, the condition in question was satisfaction of the axioms of probability, and my claim in effect was that the argument fails because the theorem is false. In the present section, the condition in question has been Reflection, and my claim has been that here the Dutch book theorem is correct but the argument based on it is fallacious. Consequently, this argument provides no reason not to draw the obvious conclusion from the counterexamples in Section 5.1.2: Reflection is not a requirement of rationality. (Christensen [1991] and Talbott [1991] arrive at the same conclusion by different reasoning.)

5.1.4 Integrity

Recognizing the implausibility of saying Reflection is a requirement of rationality, van Fraassen (1984, pp. 250–5) tried to bolster its plausibility with a voluntarist conception of personal probability judgments. He claimed that personal probability judgments 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 judgment involves a kind of commitment; to make such a judgment 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

(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 make that choice *now*, 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 you do not now endorse certain commitments, even on the supposition that you were to make them. Saying this does not undermine your status as a person of integrity.

5.1.5 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 $p(\cdot|R_q) = q(\cdot)$.[7]

Shifts that 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 $p(\cdot|R_q)$.[8] 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

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

[8] This assumes that q records everything relevant about the shift. Otherwise, it would be possible to shift from p to q in different ways (e.g., by acquiring evidence or taking drugs), some of which would involve more learning than others. This assumption could fail, e.g., if after making the shift, you would forget some relevant information about the shift. Such cases could be dealt with by replacing R_q with a proposition that specifies your probability distribution at every instant between t and $t + x$.

that 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 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.[9]

5.1.6 Reflection and rationality

Although there is nothing irrational about violating Reflection, it is often irrational to implement those potential shifts that 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 5.1.2 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, $p(H|R_q) = .5 < q(H) = 1$, and you violate Reflection; but then you would be irrational to take the drug, since the expected return from doing so is $(1/2)(\$1) - (1/2)(\$2) < 0$.[10]

[9] Jeffrey (1988, p. 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.

[10] Here and in what follows, I assume that the proviso of Section 1.9 holds. That is, I assume your probabilities and utilities are themselves rational, so that rationality requires maximizing expected utility.

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 $p(H|R_{q'}) = q'(H) = p(H) = .5$.

Assume also that

(ii) 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 probability functions that 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 the situation is fairly normal.[11]

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 that 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 $d \to R_q$ is true.

[11] According to an idea floated in Section 1.8, satisfaction of (ii) ensures that the subjective probabilities that it is rational to have are also justified.

(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 that are suitable for calculating the expected utility of the acts that will be available after the choice between d and d' is made. (What this requires is explained in the first paragraph of the proof given in Appendix A.)

(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) 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) 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 Appendix A.

The theorem can fail when the stated conditions do not hold. For one example of this, suppose you are convinced there is a superior being who gives eternal bliss to all and only those who are certain that pigs can fly. Suppose also that there is a drug that, 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.[12] 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 $p(H|R_{q_H}) = .9$, while $q_H(H) = 1$. Similarly for $q_{\bar{H}}$. Thus talking to Persi implements a shift that 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 $p(H|R_{q'}) = q'(H) = .5$. Thus not talking to Persi avoids implementing a shift that violates Reflection. Your expected return from talking to Persi is

$$(.9)(\$1) + (.1)(-\$2) = \$0.70.$$

[12] If eternal bliss includes epistemic bliss, taking the drug could even be rational from a purely epistemic point of view. Nevertheless, your certainty that pigs can fly would not be justified (see Section 1.8).

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 foregoing theorem by saying that, other things being equal, implementing a shift that violates Reflection cannot have greater expected utility than implementing a shift that 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.

5.2 Conditionalization

In Chapter 4, I noted that Bayesian confirmation theory makes use of a principle of conditionalization. The principle, as I formulated it there, was:

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 5.2.3.

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,[13] and is fallacious for the same reason.

[13] 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.

5.2.1 *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 5.1.5, 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 conditionalization is plausible only if your prior probabilities are rational; and then the subjective and objective notions of learning presumably coincide. So we can 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 that violates Reflection. But we saw, in Section 5.1.6, that there are cases in which it is rational to implement a shift that 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 that 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 5.1.6. 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 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 proved that. For we saw that when other things are equal (in a sense made precise in Section 5.1.6), 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.[14]

To summarize: The principle of conditionalization is a special case of the principle that 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 5.1.6.

5.2.2 Other arguments for conditionalization

Lewis's Dutch book argument is not the only argument that 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 that 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 prior and posterior probability functions, respectively.) Given this assumption, Teller is able to derive the principle of conditionalization. But the counterexamples that 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 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

[14]Brown (1976) gives a direct proof of a less general version of this result. What makes his result less general is that it applies only 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 5.1.2, 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 (cf. Section 5.3).

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 and Urbach 1989, pp. 67f.). Howson interprets $p(H)$ as the betting quotient on H that you now regard as fair, $p(H|E)$ as the betting quotient that you now think would be fair were you to learn E (and nothing else), and $q(H)$ as the betting quotient that you will in fact regard as fair after learning E (and nothing else). His argument is the following. (I have changed the 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 $q(H)$.

I would not endorse Howson's conception of conditional probability. But 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 differs 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.

5.2.3 Van Fraassen on conditionalization

In a recent article, van Fraassen (in press) argues that conditionalization is not a requirement of rationality. From the perspective of this chapter, 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 that will cause your posterior probability to 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 5.1.5, 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.

5.2.4 The rationality of arbitrary shifts

Speaking of the theory of subjective probability, Henry Kyburg writes:

> But the really serious problem is that there is nothing in the theory that says that a person should *change* his beliefs in response to evidence in accordance with Bayes' theorem. On the contrary, the whole thrust of the subjectivist theory is to claim that the history of the individual's beliefs is irrelevant to their rationality: all that counts at a given time is that they conform to the requirements of coherence. It is certainly not required that the person got to the state he is in by applying Bayes' theorem to the coherent degrees of belief he had in some previous state. No more, then, is it required that a rational individual pass from his present coherent state to a new coherent state by conditionalization. ... For all the subjectivist theory has to say, he may with equal justification pass from one coherent state to another by free association, reading tea-leaves, or consulting his parrot. (Kyburg 1978, pp. 176–7)

The standard Bayesian response to this objection is to claim that conditionalization has been shown to be a requirement of rationality, for example, by the diachronic Dutch book argument (Skyrms 1990b, ch. 5). But I have shown that the arguments for conditionalization are fallacious and that the principle is not a general requirement of rationality. Nevertheless, Kyburg's objection is still mistaken.

If you think there is something wrong with revising your probabilities by free association, reading tea-leaves, or consulting your parrot, then presumably shifts in probability induced by these means do not satisfy Reflection for you. If that is so, then the theorem of Section 5.1.6 shows that if these shifts would make any difference at all to your expected utility, then implementing them would not maximize expected utility, other things being equal. Thus under fairly weak conditions, Bayesian theory does imply that it is irrational for you to revise your beliefs by free association, and so forth.

5.3 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 that expresses what you learned. The principle of conditionalization is then not applicable.

Jeffrey (1965, ch. 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. But this argument 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 that 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 5.1.6.

5.4 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.