

REFLECTIONS ON REFLECTION: VAN FRAASSEN
ON BELIEF*

ABSTRACT. In 'Belief and the Will', van Fraassen employed a diachronic Dutch Book argument to support a counterintuitive principle called Reflection. There and subsequently van Fraassen has put forth Reflection as a linchpin for his views in epistemology and the philosophy of science, and for the voluntarism (first-person reports of subjective probability are undertakings of commitments) that he espouses as an alternative to descriptivism (first-person reports of subjective probability are merely self-descriptions). Christensen and others have attacked Reflection, taking it to have unpalatable consequences. We prescind from the question of the cogency of diachronic Dutch Book arguments, and focus on Reflection's proper interpretation. We argue that Reflection is not as counterintuitive as it appears – that once interpreted properly the status of the counterexamples given by Christensen and others is left open. We show also that descriptivism can make sense of Reflection, while voluntarism is not especially well suited to do so.

Van Fraassen (1984) argues for what he terms a *voluntarist* theory of avowals¹ of belief. According to such a theory, when an agent states her subjective probability for some proposition, she is undertaking an epistemic commitment. On the voluntarist view, avowals of belief have a character rather like that of performative locutions. 'I believe that *A* to degree *r*' has a character like that of 'I promise to ϕ '. With voluntarism as a theory of avowals, van Fraassen contrasts a view according to which such speech acts are reports on one's psychological state. Let us call this other view concerning avowals *descriptivism*.

Van Fraassen holds that voluntarism is true and descriptivism not. His strategy is to argue that voluntarism can, while descriptivism cannot, make sense of a principle called Reflection, which, he contends, is needed to defend traditional epistemology from a Bayesian critique. According to van Fraassen, one who wishes to understand the principles by which we infer theory from data must choose between traditional and Bayesian epistemology, and the Bayesian needs Reflection as well. So there is no reasonably comprehensive epistemology that can do

without Reflection, and voluntarism is the only account of the nature of avowals of subjective probability that can make sense of this principle.

Van Fraassen's voluntarist theory of belief has figured prominently in recent expositions of his anti-realist position in the philosophy of science.² Indeed, he claims:

[V]oluntarism . . . provides the philosophical framework . . . which I see as indispensable to the program. (1985, p. 251)

In these expositions, van Fraassen repeatedly cites 'Belief and the Will' as providing the central argument for voluntarism. An argument so central to an important philosophical position merits close scrutiny. In this paper we will challenge the claim that voluntarism is the perspective from which one can make sense of Reflection. We shall argue that descriptivism is able to account for Reflection, and that van Fraassen's claim that voluntarism is able to do so does not withstand scrutiny.

1.

According to traditional epistemology, it is possible to come to believe a scientific theory on incomplete evidence, while recognizing one's own fallibility in inferring theory from evidence. This position can be subjected to a Bayesian attack. Let H be a hypothesis or theory – say, the theory of natural selection. Let E be the statement that Bas van Fraassen will come to believe H with subjective probability 1, at some time $t + x$. Van Fraassen shows that that if, at time t , van Fraassen's subjective probability for the conjunction ($E \ \& \ \neg H$) is greater than zero, then a Dutch Book can be made against him. (A Dutch Book is a system of bets, each one deemed fair by the bettor, such that the bettor will lose money no matter the outcome.)³ If avoidance of Dutch Book vulnerability, usually called *coherence*,⁴ is a condition of rationality, it follows that it is irrational for an agent to believe that it is *possible* for her to come to be certain of a proposition that is in fact false. Van Fraassen goes on to show that we need not consider only those cases in which an agent contemplates coming to believe a proposition with certainty. Here is another case in which van Fraassen's degrees of belief leave him vulnerable to a Dutch Book:

[There will] be a race, at Hollywood Park, tomorrow at noon. The proposition H is that the horse Table Hands will run in that race and win it. The bookie now asks me seriously to consider the possibility that tomorrow morning, at 8 A.M., I shall consider fair a bet

on this proposition at odds 2 to 1. I say I do not know if that will happen – my personal probability for that eventuality, call it E , is $P(E) = 0.4$. Next he elicits my opinion about how reliable I think I am as a handicapper of horses. What is my subjective probability that E will indeed be true but that the hypothesis that Table Hands will win, is false? Suppose I answer that this degree of belief of mine, $P(\neg H \ \& \ E)$, equals 0.3. The exact numbers do not matter here too much, except that they indicate a certain lack of confidence in my own handicapping skill. In this case they entail that my present conditional probability for Table Hands winning, on the supposition that tomorrow morning I will have subjective probability $1/3$ for it, is only $1/4$. (van Fraassen, 1984, p. 241)

In general, a Dutch Book can be made against an agent a whenever her degrees of belief violate the following principle:

$$\text{REFLECTION: } P_{a,t}(A|p_{a,t+x}(A) = r) = r^5,$$

where $P_{a,t}$ is the subjective probability function of a at time t ,⁶ and ' $p_{a,t+x}(A) = r$ ' is the proposition that a 's subjective probability for A at time $t + x$ will be r . Like the diachronic principle that requires an agent to revise her degrees of belief by conditionalizing on the evidence, Reflection does not follow from the axioms of probability; it is an independent principle that must be adhered to in order to avoid the snare of the Dutch bookie.

Reflection seems to require that a rational agent have very high confidence in the accuracy of her own credence function. How is such a requirement to be understood? We are not required to have such confidence in the credence functions of others, so what sense shall we make of this injunction to hubris? Indeed, if a person anticipates drinking alcohol tonight, it would seem rational for her to *disavow* those beliefs she might form while under the influence, while paying more credence to the beliefs of others.⁷

Van Fraassen argues that Reflection may be understood as a reasonable constraint on rationality if assertions of subjective probability are seen as performatives in which the agent undertakes an epistemic commitment. 'Undertaking an epistemic commitment' in the very least involves a commitment to finding a certain wager fair: if John sincerely says 'I believe that A to degree .4', then he is committed to considering fair a bet that pays 1 if A , nothing otherwise, and which costs him .4 units of money.⁸ Promises, too, involve the undertaking of commitment, typically to later action. Further, van Fraassen suggests that promises obey an analogue of Reflection:

I say, "I promise you a horse," and you ask, "And what are the chances that you'll get

me one?" I say, "I am starting a diet today," and you ask, "And how likely is it that you won't overeat tomorrow?" In both cases, the *first* reply I must give is "You heard me!" To express anything but a full commitment to stand behind my promises and intentions, is to undermine my own status as a person of integrity and, hence, my entire activity of avowal. This applies equally in the case of conditional questions. "If you promise to marry me, will you actually do it?" "If you decide to join our crusade, will you really participate?" In the first instance these questions are not invitations to an academic discussion of the objective chances, but challenges to probe one's avowed intentions and commitments. (van Fraassen, 1984, p. 254)

By analogy, if expressing a belief amounts to undertaking an epistemic commitment, then my integrity as a judging agent requires that I stand behind my avowals of belief. Thus, if voluntarism is correct, rationality in belief-formation is analogous to sincerity in making promises, and we should expect a condition like Reflection to be necessary for rationality.

Van Fraassen argues further that if avowals of subjective probability were merely autobiographical reports – that is, if descriptivism were true and voluntarism false – then Reflection would remain mysterious, for there would be no justification for believing our own psychological states to be more reliable than those of another.

In the final section of 'Belief and the Will', van Fraassen claims that traditional epistemology remains immune from attack as long as the epistemic agent is required to adhere to Reflection. This is a bit premature, for there is still the worry that if an agent updates her personal probabilities in a way that deviates from simple conditionalization on the evidence, then she will be vulnerable to a diachronic Dutch Book.⁹ Because traditional epistemology, as characterized by van Fraassen, seems to involve leaps of faith that go beyond mere conditionalization, adherence to Reflection is still insufficient to shield an agent from the wiles of the Dutch bookie. The missing piece is supplied by van Fraassen (1989).¹⁰ There he argues that the diachronic Dutch Book argument shows only that if one updates one's subjective probabilities *according to some rule*, that rule must be equivalent to conditionalization on the evidence. Spontaneous changes of degrees of belief do not leave the agent vulnerable to the Dutch bookie, who can only concoct the Dutch Book if he knows in what way the agent's belief revisions will deviate from conditionalization.¹¹ Traditional epistemology, with its leaps of faith, is unscathed, as long as one's inductive leaps do not follow a prescribed rule (other than conditionalization), and as long as they adhere to Reflection.

2.

In this section, we shall make two exegetical digressions. The first concerns the very point of propounding voluntarism. Van Fraassen talks of voluntarism as providing a defense or justification of Reflection; e.g.:

... I propose to examine, and indeed advocate, serious attempts to defend the principle. (p. 244)

[J]ustification of this principle can follow no ordinary route. (Ibid., p. 245)

The principle (Reflection) can be defended ... if we give a different, voluntarist interpretation of epistemic judgment (p. 256)

This is curious; if Reflection is forced on us by the Dutch Book argument, why is it in need of defense? The difficulty may be made vivid by presenting van Fraassen with a dilemma. Is the diachronic Dutch Book argument for accepting Reflection as a canon of rationality cogent?¹² If so, then no other defense is needed. If not, then we are not compelled to search for a defense; we are free to reject Reflection in light of its admittedly unintuitive character.

It seems to us that the best way through the horns of the dilemma is to reconsider the relationship between voluntarism and Reflection. An analogy is provided by the relationship that holds between theory and phenomena in science. *Ceteris paribus*, a good theory should be able to accommodate an observed phenomenon. What is meant by the accommodation of phenomena by theory? Ideally, the theory would, together with a description of initial conditions, *entail* a description of the phenomenon, although weaker forms of accommodation can be countenanced. Two features of the relation between theory and phenomena should draw our attention. First, the point of accommodation is *not* to lend whatever credibility the theory might have to the phenomenon. Second, phenomena are typically inferred from experimental data; a theory should accommodate a phenomenon independently of such an inference. In other words, the entailment of a phenomenon by a theory should not retrace the steps of the inference by which the phenomenon was inferred from data. Analogously, if the Dutch Book argument shows that Reflection must be accepted as a principle of rational belief, it is a desideratum of any theory of belief that it be able to accommodate this principle independently of the Dutch Book

considerations which led to it. Voluntarism, van Fraassen suggests, independently entails this principle, whereas descriptivism strongly suggests its denial.

We turn now to the second exegetical digression. Just what does Reflection demand of rational epistemic agents? Van Fraassen sometimes suggests that Reflection requires an agent to have perfect confidence in her own future assessments of chances; but this informal gloss is misleading. In telling his story of the race at Hollywood Park, for example, van Fraassen describes his violation of Reflection in the following way:

The . . . numbers . . . indicate a certain lack of confidence in my own handicapping skill.

This type of language suggests a principle much stronger than Reflection, something like:

$$\text{CONFIDENCE: } P_{a,t}(p_{a,t+x}(A) \neq ch_{t+x}(A)) = 0;$$

or one of its consequences:

$$\text{CONFIDENCE': } P_{a,t}(p_{a,t+x}(A) = r | ch_{t+x}(A) = r) = 1,$$

or

$$\text{CONFIDENCE'': } P_{a,t}(ch_{t+x}(A) = r | p_{a,t+x}(A) = r) = 1,$$

for all r for which the conditional probabilities are defined. Here ' $ch_{t+x}(A) = r$ ' is the proposition that the objective chance of A at time $t + x$ is r .¹³ The Confidence Principle, then, says that the rational agent should be certain that her degree of belief in the proposition A at time $t + x$ will be a correct estimate of the chance that A . This seems a more plausible construal of what it would be for an agent to have no "lack of confidence in [her] own handicapping skill".

Reflection can hold when Confidence fails. The exposition will be aided if we introduce the 'Principal Principle', advocated by David Lewis (1980):

$$\text{PRINCIPAL: } P_{a,t}(A | ch_{t+x}(A) = r \ \& \ E) = r,$$

for admissible E .¹⁴

The proposition E is said to be admissible if it does not smuggle in any information about whether A does in fact hold, that is not contained in the statement ' $ch_{t+x}(A) = r$ '. A sufficient condition for E to be

admissible is that it contain only historical information about goings-on up to and including time $t + x$.¹⁵

An example will suffice to show that Reflection can hold when Confidence fails.¹⁶ In this example we have $x = 0$. A box contains two coins, one that the agent a believes to be fair, and one that she believes to be biased so as to have a 2 in 3 chance of landing heads. She picks a coin at random and flips: What is her probability for heads? Since a is equally likely to pick either coin, we have

$$(1) \quad P(ch(H) = 1/2) = 1/2; P(ch(H) = 2/3) = 1/2.¹⁷$$

Applying Principal yields

$$(2) \quad P(H|ch(H) = r) = r.$$

By (1), (2), and the theorem of total probability,

$$(3) \quad P(H) = (1/2)P(H|ch(H) = 1/2) + (1/2)P(H|ch(H) = 2/3) \\ = (1/2)(1/2) + (1/2)(2/3) = 7/12.$$

If a knows that $P(H) = 7/12$, then $P(p(H) = 7/12) = 1$. Since conditioning on a set of measure 1 does not alter probability,

$$(4) \quad P(H|p(H) = 7/12) = 7/12.$$

Thus Reflection is satisfied. But (1) entails that $P(ch(H) \neq 7/12) = 1$, so $P(ch(H) \neq p(H)) = 1$, in violation of Confidence.¹⁸ Because the agent's personal probability of heads is the weighted average of (what she believes to be) the two possible objective chances, she does not believe that her personal probability will agree with the *actual* objective chance. Thus Reflection does not require that the agent believe herself to estimate objective chances accurately.

What, then, does Reflection require of the rational epistemic agent? In the appendix, we prove that relative to Principal, Reflection is equivalent to

$$\text{BALANCE: } \sum_t t \times Pr \text{ (overestimating the chance of } A \text{ by } t) \\ = \sum_t t \times Pr \text{ (underestimating the chance of } A \text{ by } t).¹⁹$$

Reflection is equivalent to a principle that, loosely put, requires that an agent believe herself to be no more likely to overestimate the chance of an event than to underestimate it. (This formulation is loose in that it ignores the weighting of the probability of under- or overestimating

by the amount of the under- or overestimation.) The mathematical derivation exploits Lewis's Principal Principle, which is not beyond reproach. Nonetheless, it serves to motivate a new informal interpretation of Reflection that plausibly describes the beliefs of the reflective agent.

Return to the horse race at Hollywood Park. In van Fraassen's story,

$$P_{a,t}(p_{a,t+x}(H) = 1/3) = .4, \text{ and}$$

$$P_{a,t}(p_{a,t+x}(H) = 1/3 \ \& \ \neg H) = .3,$$

where t is today, $t + x$ is tomorrow at 8:00 A.M., and H is the proposition that Table Hands will win the race at noon. The numbers given above entail that

$$P_{a,t}(H|p_{a,t+x}(H) = 1/3) = 1/4,$$

in violation of Reflection. Van Fraassen describes the situation as one in which the numbers represent a lack of confidence in his handicapping skill. But the numbers do more than that: they indicate van Fraassen's belief that he is more likely to *overestimate* Table Hands' chances of winning than to *underestimate* them (on the hypothesis that he finds 2 to 1 odds fair tomorrow at 8:00 A.M.). There is a lopsidedness in van Fraassen's lack of confidence. It is this lopsidedness, rather than his lack of confidence per se, that constitutes the violation of Reflection in van Fraassen's degrees of belief. This understanding of Reflection should prove troublesome to van Fraassen for two reasons: first, because it suggests that Reflection is weaker, and so less counterintuitive, than first appeared, there is less call for a non-standard account of avowals of subjective probability; second, because it is possible for an agent to have no confidence in the accuracy of her estimates of chance whatsoever and still be 'balanced', it does not seem that such an agent is best described as 'standing behind' her beliefs. We will explore these points in more detail in the next three sections.

3.

With a better informal understanding of what Reflection demands of the rational agent, we can approach the problem of trying to give an intuitive account of why degrees of belief that violate Reflection are irrational. In this section we will also explore prospects for a descriptivist accommodation of Reflection.

What would a descriptivist theory of personal probability avowals look like? A first attempt might be as follows: when an agent a avows that $P_{a,t}(A) = r$, she is describing a psychological state that she has at time t . This formulation is subject to a straightforward objection. Actual agents are prone to epistemic inadequacies, and thus are unlikely to have degrees of belief that conform precisely to the probability calculus. Thus, when an agent describes herself as having a certain subjective probability for A , the description is either false, or employing a sense of ‘probability’ that is not moored to the standard probability axioms. The psychological states of the typical agent simply cannot be described using the probability calculus. Van Fraassen, however, does not offer this objection to descriptivism, and this suggests that he has a more sophisticated descriptivist theory as his target.

Van Fraassen has referred to the probability calculus as the “logic of judgment” (e.g., in 1989, pp. 155–59), suggesting that the probability calculus provides a normative account of beliefs. But the probability calculus could simultaneously serve both a descriptive and a normative role. Describing the early history of probability theory in the Age of Reason, Gigerenzer et al. (1989, p. 16) write:

Only a small élite of *hommes éclairés* could reason accurately enough by unaided intuition, the calculus of probabilities sought to codify these intuitions . . . for use by *hoi polloi* not so well endowed by nature.²⁰

One may construe probability theory as, from its inception, playing a normative role by *describing* the judgments of a rational role model. Such a role model could legitimately use the probability calculus to describe her own states of partial belief. Likewise, we may also construe avowals of subjective probability as made by the *hoi polloi* as autobiographical reports on psychological states that would, in the ideal limit, conform to the probability calculus.

That van Fraassen has such a formulation of descriptivism in mind is suggested by his response to a remark by David Lewis that might be used to provide an account of Reflection:

A . . . possibility was advocated in discussion by David Lewis: that the standard of rationality exemplified by Dutch Book Invulnerability applies to a certain sort of ideally rational agent, who not only believes himself to be, but is infallible with respect to perception, and which we explicitly realize ourselves not to be. But this leaves us still with the task of constructing an epistemological theory that does apply to our own case. (van Fraassen, 1984, p. 243, n. 10)

A proposal by Sobel (1987) develops Lewis's suggestion in more detail. Sobel attempts to explain how an ideal intellect may be expected to adhere to Reflection, in spite of the *prima facie* implausibility of that principle. Sobel (1987, pp. 68–69) gives a set of features that an ideal intellect must have:

1. *consistency*: he does not believe a set of propositions not all of which can be true.
 2. *balance*:²¹ his opinions are represented only by additive credence functions.
 3. *logical omniscience*: he is certain of every necessary truth.
 4. *high opinionation*: he assigns sharp values to his degrees of belief in all propositions.
 5. *introspective omniscience*: he is certain of, and knows, the state of some of his opinions.
- The opinions in question are of three kinds:
- (i) his concurrent opinions
 - (ii) his past opinions
 - (iii) some of his future opinions, namely those opinions he will have to the effect that he will be rational in the sense of 1 through 5.

Sobel argues that anyone who adheres to these five criteria is also Reflective. Now, Sobel's ideal, surely in the spirit of Lewis,²² is a demanding one, but it will not do to reject it by remarking that such an ideal is unattainable. The reason is that it is quite difficult to assess such a claim, ignorant as we are of absolute human limitations. On the other hand, we suggest that a fair response for van Fraassen is to maintain that Sobel's idealization is so great as to justify hope for a more accurate model, in which are represented various of the imperfections and failings that we know to apply to our own case. *Ceteris paribus*, the less idealization the better, and one has a right to hope for something less idealized than what Lewis and Sobel offer.

A better tack seems to us to be the following. The calculus of probabilities should be thought of as describing the belief states of an agent who represents a more proximate ideal – we will call him Rational Joe. Rational Joe does not assign a sharp degree of belief to every proposition in his language; indeed, there are some strings of symbols so complex that he cannot determine whether they are well-formed propositions. A probability function defined on propositions is said to represent Joe's set of beliefs if it respects his judgments of likelihood. For example, if among his beliefs we find 'It seems to me that it is more likely to rain than snow tomorrow', then any probability function representing Joe's beliefs must assign a higher probability for the proposition 'rain tomorrow' than for the proposition 'snow tomorrow'. Synchronic Coherence requires that the set of all Joe's beliefs can be consistently

represented by some set of probability functions.²³ Rational Joe avoids obvious pitfalls, such as offering 2 to 1 odds on both outcomes of a coin toss. In situations in which he lacks sufficient information, he prudently refrains from assigning sharp probability values. In short, Rational Joe provides us with an epistemic role model without being either omniscient or infallible. The question is now: Do Rational Joe's belief states obey Reflection?

Suppose that Rational Joe's personal degrees of belief do not obey Reflection. Suppose, for example, that Rational Joe has the degrees of belief described by van Fraassen in the Hollywood Park example; thus Rational Joe believes that if he will come to have degree of belief $1/3$ for the proposition 'Table Hands will win the race at noon', he will be more likely to have overestimated Table Hands' chances of winning than to have underestimated them. This suggests that either (i) Rational Joe believes himself to have a habit of overestimating chances, and that he believes he will persist in this habit, or (ii) Joe believes some special circumstance will cause him to overestimate chances. In either case, despite recognizing his proclivity to overestimate the chances in question, he has not (or at any rate, does not believe himself to have) adopted a policy of correcting for this tendency. Joe has given himself up to a kind of intellectual abandon, and we may view Reflection as an admonition against thus letting oneself go. This, we claim, is a natural constraint to impose on a rational belief-forming agent. It does not put an undue burden on Rational Joe's cognitive capacities, for it only forbids him to leave unchecked those tendencies to over- and underestimate chances that he explicitly recognizes himself to have.²⁴

The foregoing is consistent with Rational Joe's avowals being autobiographical reports, rather than undertakings of commitment. Joe's being Reflective amounts, as we have suggested, to his being just as likely to overestimate as to underestimate chances, whence those of his avowals that conform to Reflection may be seen as describing himself as being balanced. Further, Joe corrects those of his tendencies to over- or underestimate chances that he recognizes himself to have, but it is a far cry from this to the view that Joe in general corrects those tendencies to over- or underestimate that he recognizes in others. The descriptivist may use this asymmetry to capture the fact that Joe is Reflective but need not be in epistemic solidarity with others. That is, as Christensen (1991) points out, an implausible generalization of Reflection is the principle of

$$\text{SOLIDARITY: } P_{a,t}(A|p_{b,t+x}(A) = r) = r,$$

where a and b name distinct individuals. According to the description of Rational Joe sketched above, Joe is Reflective but not taken to be in Solidarity with others.

This picture may render Reflection plausible. But that does not render this putative norm immune to some counterexamples that have recently been lodged against it. In an attempt to show that Reflection is no sort of rational ideal, Christensen (1991) offers cases in which it is rational to be irreflective. He writes,

[Reflection] is more than just initially implausible; it is wrong. In some cases, violations of Reflection are not only rationally permissible, but mandatory; to respect Reflection in such cases would itself constitute a grave intellectual imperfection. To put it bluntly, there are cases in which satisfying the principle of Reflection would be downright stupid. (Christensen, 1991, p. 230)

Suppose there is a drug called LSQ that causes those who ingest it to become convinced of their ability to fly by flapping their arms. Suppose Mary knows that she has just swallowed a dose of LSQ. Would it not be rational for Mary to violate Reflection by saying, ‘The probability that I’ll be able to fly in one hour, given that I’ll then take the probability that I can fly to be .99, is still very low’? This is the rational thing for Mary to hold, Christensen points out, but it is also in outright violation of Reflection. To the objection that anyone who takes LSQ is irrational in the first place, Christensen rightly responds that this need not be so: there might be strong overriding reasons for ingesting the drug, or Mary may have unwittingly ingested it, thinking it was Kool Aid.²⁵ Indeed, Mary may be told that what she drank was LSQ, though it in fact was Kool Aid, and it would still be rational for her to be irreflective.

One might still feel that the reasonableness of Mary’s being irreflective in such cases turns on her *envisioning* a later lapse from rationality. But there are examples in which it is rational to be irreflective where the agent does not envision any lapse from rationality. Talbott (1991) provides the following example. Let ‘ S ’ abbreviate ‘Mary has spaghetti for dinner today’, let ‘ t ’ name today, and let ‘ x ’ name a span of one year. Suppose further that she generally eats spaghetti for dinner about one time in ten, and that she knows this fact about herself. She is also quite confident that today she will have spaghetti for dinner, for she

has made up her mind to do so. Hence, she may be described by the following formula:

$$(i) \quad P_t(S) = 1.$$

A year from now she is quite unlikely to recall what she had for dinner today, and she now takes it that she will feel, one year from now, that there is one chance in ten that she had spaghetti for dinner today. But it would still seem rational for her to conform to

$$(ii) \quad P_t(S|p_{t+x}(S) = .1) = 1$$

and not

$$(iii) \quad P_t(S|p_{t+x}(S) = .1) = .1.$$

We have seen two cases in which it appears rational to be irreflective. Now, one should keep one's promises, but in certain cases it is best to break them. One should tell the truth, but in certain cases it is best to lie. There are many norms that, in a given circumstance, it is reasonable to violate, but this fact need not vitiate such norms. Moreover, in the case of epistemic norms, added confusion is caused by the labile word 'rational', which can be applied to prudential as well as epistemic norms.²⁶ If a wealthy benefactor were to offer van Fraassen one million dollars on the condition that he form an irreflective belief, few would doubt that it would be rational for him to do so. Indeed, for such a payoff, it may well be rational to affirm the consequent, beg the question, and commit all manner of heinous fallacies. 'Rational' is here being used in the prudential, rather than the epistemic, sense; the moral is that the existence of cases in which it is prudent to violate an epistemic norm does not vitiate the norm.²⁷

There are, however, other cases in which epistemic rationality, not just prudence, dictates violation of a putative epistemic norm, and this kind of situation may appear to be a genuine threat to the viability of that putative norm. For example, Clifford's (1879) injunction never to believe any proposition on insufficient evidence seems to be vitiated by the fact that one who wishes to understand nature will be seriously hindered unless she can believe things on insufficient evidence. We are not, however, obliged to eschew a putative epistemic norm in response to cases in which it is epistemically rational to violate that putative norm. We suggest Ross's (1930) understanding of ethical principles as a paradigm for how we may conceive of principles of epistemic rational-

ity: ethical principles are at best *prima facie*, and, further, although there are clear cases in which it is morally permissible or obligatory to violate such principles, we should not insist that the conditions under which these principles do not apply admit of exhaustive specification. On Ross's approach it is not the job of the ethical theorist to provide an algorithm for making the moral choice in all possible situations. Similarly, we should deny that it is the job of the epistemologist to provide an algorithm for rational belief-formation in all possible situations. This is in contrast to the way in which the decision theorist purports to provide a recipe for rational action in all possible choice situations.

Note well that it is quite consistent for Ross to attempt to sharpen and streamline various of the *prima facie* norms. Likewise, it would be consistent for us to attempt a rough formulation of the conditions under which a principle such as Reflection may be expected not to apply. A contribution to this end is made by Maher (1992). While rejecting Reflection as a canon of rationality, Maher proves that, subject to five conditions, implementing those beliefs about which one is reflective²⁸ is always rational in the sense of maximizing expected utility.²⁹ The five conditions enumerated by Maher are clearly candidates for being *ceteris paribus* conditions of the sort desired.

We cannot hope that the foregoing line of response will satisfy all who would deny Reflection on the basis of examples such as those we have discussed. But we have made a case for the reasonableness of a position that supports Reflection in spite of such examples (without saying that this is the only reasonable line). This, further, is enough to show that the presence of such examples does not render nugatory questions about the relations among Reflection, descriptivism, and voluntarism. Indeed, a case can be made that an evaluation of the intuitions that underlie these counterexamples cannot be made independently of a theory about the nature of belief.

4.

We turn to the question of whether van Fraassen's claim that voluntarism can account for Reflection withstands scrutiny. Van Fraassen makes only impressionistic remarks about how voluntarism is to be understood, and we are left wondering how far, for example, we are to take the analogy between avowals and promises. Under the right conditions

one who says 'I promise to ϕ ' thereby promises to ϕ ; should we take it that on van Fraassen's view one who, under the right conditions, says 'I believe that A to degree r ' thereby believes A to degree r ? One must tread carefully here, especially because the name 'voluntarism' can mislead. Voluntarism is not the view that we can form beliefs at will: 'I pronounce you man and wife' is a performative, but it does not follow that anyone can pronounce anyone man and wife just by uttering these words.

The aspect of the analogy with promising that we shall emphasize is the fact that one who promises *undertakes a commitment*. The commitment involved in a promise is to later action. The commitment has the effect of placing an evaluative framework on the future actions of the agent: those that fulfill the commitment are appropriate, while those that do not, are not. Moreover, the evaluative framework applies indirectly to others of the agent's promises: it is inappropriate, for instance, to promise to perform an action that one is already committed to refraining from. Similarly, we shall take it that according to voluntarism one who avows a subjective probability undertakes a commitment, which places an evaluative framework over the agent's actions and other degrees of belief. In this instance, those actions and beliefs that conform to the commitment are deemed rational. One part of the commitment undertaken is to finding a certain wager fair. One who says 'I believe A to degree r ' is committed to finding fair a wager that pays one unit of money if A , nothing if $\neg A$, and which costs r units of money – call this the *r-wager on A*. Being committed to finding a wager fair does not imply being committed to purchasing that wager, as the critics of the Dutch Book arguments are fond of pointing out.

Van Fraassen seems to hold that this picture of avowals of personal probability implies Reflection:

[M]y integrity, qua judging agent, requires that, if I am presently asked to express my opinion about whether A will come true, on the supposition that I will think it likely tomorrow morning, I must stand by my own cognitive engagement as much as I must stand by my own expressions of commitment of any sort. I can rationally and objectively discuss the possibility of a discrepancy between objective chance and my previsions. But I can no more say that I regard A as unlikely on the supposition that tomorrow morning I shall express my high expectation of A , than I can today make the same statement on the supposition that tomorrow morning I shall promise to bring it about that A . To do so would mean that I am now less than fully committed (a) to giving due regard to the felicity conditions for this act, or (b) to standing by the commitments I shall overtly enter. (1984, p. 255)

The suggestion is that if voluntarism is true, then a person who has integrity as a judging agent must respect Reflection as a norm to which to aspire. Hence, voluntarism is supposed to imply Reflection, and thereby make that principle intelligible. Van Fraassen must here be understood as claiming that voluntarism provides an independent motivation for Reflection. If voluntarism only entails Reflection in conjunction with Dutch Book considerations, then voluntarism does not provide us with anything that we do not already have.

In defending this implication claim, van Fraassen appeals to the analogy with promising and other performatives, although the promisor's analogue of Reflection is never explicitly stated. The analogue may be readily reconstructed, however. For it would seem to be a condition of one's being capable of sincere undertakings of commitments to act, that that person be willing to affirm that, if she commits herself to doing ϕ , then she will do ϕ . To elucidate this idea, let us introduce the notion of a conditional promise. One such example would be,

CP: I promise to buy you a new car, provided that you get all A's on your report card.

We must take care to distinguish CP from a conditional with a promise described in the consequent, such as:

CPC: If you get all A's on your report card, then I will promise to buy you a new car.

In CPC, it is said that a promise will be made if the antecedent obtains; CPC is not a promise at all.³⁰ In CP, a promise is made at the time of utterance, although the commitment to buy a car does not arise until the antecedent is satisfied. Let us call this commitment the *major commitment* of CP. The major commitment of a conditional promise arises if and when the condition stated in the promise obtains. There is also, however, a *minor commitment* involved in CP that comes into force at the time the promise is made: this commitment is, roughly, to keep open the possibility of purchasing a new car, until and unless it becomes obvious that the promisee will not bring home a report card with solid A's. A parent who issues CP and then spends all of her savings on a Persian rug is acting in bad faith.³¹

A special case of a conditional promise is one whose condition de-

scribes a promise – let us call such a promise a ‘reflective promise’. An example would be:

RP: I hereby promise that, if tomorrow I promise to buy you a horse, then I will buy you a horse.

Note that, while the condition expresses a promise made tomorrow, the conditional promise is made today. In particular, RP is to be distinguished from the trivial conditional:

If tomorrow I promise to buy you a horse, then I will promise to buy you a horse.

It seems natural to expect a person in the habit of making promises to be willing to make reflective promises. Our being aware of the fact that tomorrow Jane may be intoxicated or under duress seems not to detract from our expectation that she be willing to make such a conditional promise, for the promise hypothesized in the antecedent is somehow assumed to be made under normal circumstances. If Jane were explicitly to refuse to make a promise such as RP, we would naturally regard her as disavowing her own integrity as a promising agent. It is possible to understand the requirement that sincere promisors be willing to make reflective promises in terms of the major and minor commitments of such promises. The major commitment of RP, the commitment to buy a horse, arises if and when the unconditional promise to buy a horse is made. A sincere promisor should be willing to commit herself to following through on her *unconditional* promises in this way. The minor commitment is simply to avoid making some subsequent promise that one is in no position to keep; the commitment can be fulfilled either by leaving open the option of buying a horse, or by leaving open the option of not making the promise. This is a weak commitment, and certainly no more than one would expect of a potential promisor. Uttering RP does not bring about any new commitments; it simply makes explicit the commitments that already apply to one as a sincere maker of unconditional promises.

Willingness to make reflective promises is such a natural thing to expect of a person in the habit of making promises that we may lionize it in the form of a

PROMISOR’S PRINCIPLE: One should, at time *t*, be wil-

ling to promise to ϕ (at time $t + x$), conditional on one's promising, at time $t + x$, to ϕ .

Evidently, one can generalize the principle here to other cases in which a person undertakes a commitment, with a

SINCERITY PRINCIPLE: One should, at time t , be willing to undertake a commitment to ϕ (at time $t + x$), conditional on one's undertaking a commitment, at time $t + x$, to ϕ .

To prefer A to B is, among other things, to be committed to choosing A over B , all else being equal. Thus the Sincerity Principle gives us,

I hereby commit myself to (tomorrow) choosing A over B (if given the opportunity), provided that tomorrow I prefer A to B .

And for probability assignments, on the assumption that believing a proposition A with probability r commits one to finding fair the r -wager on A :

ANALOGUE: I hereby commit myself, given that tomorrow I believe A with strength r , to (tomorrow) taking r to be a fair price for a wager that pays 1 if A and nothing otherwise.

Analogue is supported by a voluntarist conception of avowals of personal probability. Moreover, Analogue looks like an informal statement of Reflection; but is it? Consider the commitments that arise as a result of Analogue, and the times at which those commitments come into force. The major commitment of Analogue is to finding the r -wager on A fair tomorrow, on the condition that $P_{a,t+x}(A) = r$. The minor commitment, which is in effect today, is to avoid the conjunction of (i) $P_{a,t+x}(A) = r$ and (ii) the agent's not finding the r -wager on A fair tomorrow. Compare these to the commitments to which

REFLECTION: $P_{a,t}(A|p_{a,t+x}(A) = r) = r$

gives rise. It shares the major commitment of Analogue, namely, to find fair the r -wager on A tomorrow if $P_{a,t+x}(A) = r$. Reflection also involves a commitment that is in effect at time t , namely to finding fair a *conditional r -wager on A* , the condition being that $P_{a,t+x}(A) = r$.³²

This commitment seems to be much stronger than the minor commitment associated with Analogue: if there is to be an argument for Reflection from voluntarism via Analogue, the minor commitment of Analogue would have to entail the commitment that is undertaken at time t by the adherent of Reflection.

Here is an attempt to establish this entailment. In virtue of Analogue, the agent who is 'epistemically sincere' must keep open the option of finding fair the r -wager on A at time $t + x$, should it turn out that $P_{a,t+x}(A) = r$. Suppose that the agent has already, at time t , taken out a *conditional* s -wager on A , on the condition that $P_{a,t+x}(A) = r$, with $s > r$.³³ Then there is a sense in which she has precluded the possibility of finding the r -wager on A fair tomorrow, should $P_{a,t+x}(A) = r$; for if this circumstance were to obtain, and she were to offer the r -wager on A to the bookie, she could be guaranteed of suffering a loss.

The problem with this argument is that it devolves on the Dutch Book framework: the agent's commitment to keep open the option of finding certain bets fair in the future entails her commitment to finding the earlier, conditional bet fair only if she is also committed to avoiding Dutch Book. Thus, the argument does not provide a derivation of Reflection from voluntarist principles alone: we do not have the independent motivation for Reflection that was promised.

Our argument that voluntarism does not entail Reflection is not knock-down, for we have examined the prospects for such a derivation only from within a sketchy formulation of a voluntaristic theory of belief. In particular, the only commitments we have associated with avowals of personal probability are commitments to finding certain wagers fair. Perhaps by employing the resources of a full-blown voluntarist theory of belief the entailment can be pushed through. (We shall consider a different attempt in the next section.) Our sketchy version, however, uncovers an obstacle that must be overcome. The Sincerity Principle, which enjoins agents to 'stand behind' their future commitments, enjoins agents to commit themselves to avoiding future circumstances in which they will make unfulfillable commitments. In particular, it enjoins them not to perform actions or to make commitments that will be incompatible with later commitments. When applied to avowals of personal probability, it enjoins an agent not to avow personal probabilities that will be incompatible with any future degrees of belief she might avow. But there's the rub: What is it for present

and future degrees of belief to be incompatible? The challenge is to find an answer to this question that does not fall back on Dutch Book arguments.

5.

In the previous section, we assumed that the commitment undertaken by the avowal ' $P_{a,t}(A) = r$ ' was a full commitment to finding the r -wager on A fair at t . It would be natural, however, to assume that, on a voluntarist theory of personal probability avowals, many of the commitments involved will be partial. Thus the commitment undertaken by the avowal ' $P_{a,t}(A) = r$ ' would be to the truth of A , with strength r . In this section, we show that the voluntarist's argument for Reflection is not aided by this incorporation of the notion of graded commitment.

How are we to make sense of graded commitment? Speech acts of the same general type may differ in strength.³⁴ A directive speech act, for example, may be a command, a request, an imploring, or a suggestion. Performatives, too, may vary in the strength of commitment undertaken. Compare:

As God is my witness, I *will* avenge my father's death!³⁵
 I promise to buy you a horse.
 I will try to make it to your party.
 Perhaps I will stop by, if I can get away.

A promise that is less than fully committal might be called a halfhearted promise – let us generalize this to the notion of an r -hearted promise, where $0 \leq r \leq 1$. An agent who utters an r -hearted promise undertakes a commitment of strength r . But do not let the language mislead: an r -hearted promise is not, as the name might suggest, a promise that is less than sincere.

The argument of the previous section showed that voluntarism entails the Sincerity Principle. That discussion, however, did not concern graded commitments. The natural interpretation of that argument takes it to have established that the Sincerity Principle holds for *wholehearted* commitments. Let us restate the Sincerity principle so as to make this explicit:

SINCERITY: One should, at time t , be willing to undertake

a wholehearted commitment to ϕ (at time $t + x$), conditional on one's undertaking a wholehearted commitment to ϕ at time $t + x$.

Leaving aside the worries of the previous section, let us consider the following proposal: understand ' $P(A) = r$ ' as the undertaking of an r -hearted commitment to the truth of A , and ' $P(A|B) = s$ ' as an s -hearted conditional commitment. Then Sincerity gives us

$$\text{SPECIAL REFLECTION: } P_{a,t}(A|p_{a,t+x}(A) = 1) = 1,$$

which is a special case of Reflection. It is not clear, however, that Reflection is the appropriate generalization of Special Reflection. As we show in the Appendix, Special Reflection is equivalent, relative to Principal, to a special case of Confidence":

$$\text{SPECIAL CONFIDENCE"}: P_{a,t}(ch_{t+x}(A) = 1|p_{a,t+x}(A) = 1) = 1.$$

So perhaps Confidence" is the appropriate analogue of the Sincerity Principle for avowals of personal probability. If so, then voluntarism is surely the wrong theory of belief, for Confidence" is too strong to be a constraint on rationality.

The problem reduces to one of how to generalize Sincerity to cover commitments with strengths less than one. There are at least two candidates:

REFLECTIVE SINCERITY: One should, at time t , be willing to undertake an r -hearted commitment to ϕ (at time $t + x$), conditional on one's undertaking an r -hearted commitment to ϕ at time $t + x$.

CONFIDENT SINCERITY: One should, at time t , be willing to undertake a wholehearted commitment to ϕ 'ing (at time $t + x$) *with chance* r , conditional on one's undertaking an r -hearted commitment to ϕ at time $t + x$.

Our intuitions are not refined enough to help us to prefer one over the other, or indeed over other possible generalizations of Sincerity. We do have a clue, however: it was shown in Section 2 that one could adhere to Reflection without having any confidence in the reliability of one's own estimates. This suggests that adhering to Reflection is not much like standing behind one's own commitments, which in turn

suggests that Reflection is not the appropriate generalization of Sincerity.

6.

Contrary to what Austin once held, 'I promise to ϕ ' can at once be a performative and have truth value. There are no good arguments from the performative character of a locution to its lack of a constative or descriptive dimension, and there are plenty of reasons for taking the surface indicative character of many performative locutions as entitling them to truth value.³⁶ In 'Belief and the Will', and even more explicitly in *Laws and Symmetry*, van Fraassen insists on a distinction between avowals and ascriptions in such a way as to suggest that a given speech act cannot be both an avowal and an ascription of an attitude. We see no obstacles to taking 'I believe that A ' as, in some cases, undertaking a commitment, *and* as describing the agent as being in a certain psychological state. In the same way, 'I promise' describes the speaker as promising *and*, typically, is the undertaking of a commitment to act.

Van Fraassen claimed that the view that Reflection is a canon of rationality is implausible from the perspective of a descriptivist theory of belief, but that it is precisely what one would expect from the perspective of voluntarism. We have argued against both of these claims. It is still an open question whether avowals must be seen as having a voluntarist dimension, for we surely have not proven that they do not. Should this question be answered in the affirmative, however, more will be required before we are justified in taking avowals to have no descriptive dimension.

APPENDIX

1.

Assume

PRINCIPAL: $P_{a,t}(A|ch_{t+x}(A) = r \ \& \ E) = r$ for admissible E .

We will show that

REFLECTION: $P_{a,t}(A|p_{a,t+x}(A) = r) = r$

is equivalent to

BALANCE: $\sum_t t \times Pr$ (overestimating the chance of A by t) = $\sum_t t \times Pr$ (underestimating the chance of A by t).

For a fixed proposition A , the conditional probability $P_{a,t}(\cdot|p_{a,t+x}(A) = r)$ is a probability measure (if it is defined at all); let us abbreviate it by Pr . That is, for any proposition B in the relevant algebra of propositions, $Pr(B) = P_{a,t}(B|p_{a,t+x}(A) = r)$. Reflection then requires that $Pr(A) = r$. Using the theorem of total probability, this can be rewritten as:

$$\sum_s Pr(A|ch_{t+x}(A) = s) \times Pr(ch_{t+x}(A) = s) = r;^{37}$$

i.e.,

$$\begin{aligned} \sum_s P_{a,t}(A|ch_{t+x}(A) = s \ \& \ p_{a,t+x}(A) = r) \\ \times P_{a,t}(ch_{t+x}(A) = s|p_{a,t+x}(A) = r) = r. \end{aligned}$$

By Principal, the left-hand factor collapses to s .³⁸ So,

$$(*) \quad \sum_s s \times P_{a,t}(ch_{t+x}(A) = s|p_{a,t+x}(A) = r) = r.^{39}$$

For $r \neq s$, the right-hand factor may still be non-zero, although it cannot be one. Note that $\sum_s Pr(ch_{t+x}(A) = s) = 1$. Multiply both sides of this equation by r , and subtract each side from the corresponding side of (*), and the result is

$$\sum_s (s - r) \times Pr(ch_{t+x}(A) = s) = 0,$$

which can be rewritten as:

$$\begin{aligned} \sum_{s>r} (s - r) \times Pr(ch_{t+x}(A) = s) = \\ \sum_{s<r} (r - s) \times Pr(ch_{t+x}(A) = s); \end{aligned}$$

which, when put less formally, gives us Balance. The proof holds for every A and r for which $P(\cdot|p_{a,t+x}(A) = r)$ is defined. Because every step in the proof is reversible, Reflection and Balance are equivalent.

2.

We now show that, assuming Principal,

$$\text{SPECIAL REFLECTION (SR): } P_{a,t}(A|p_{a,t+x}(A) = 1) = 1$$

is equivalent to

$$\text{SPECIAL CONFIDENCE}'' \text{ (SC): } P_{a,t}(ch_{t+x}(A) = 1 | p_{a,t+x}(A) = 1) = 1.$$

Let $Pr(\cdot) = P_{a,t}(\cdot | p_{a,t+x}(A) = 1)$. Both SR and SC imply that the conditional probability is well defined. Assume SR. Now if $Pr(ch_{t+x}(A) < 1) > 0$, we can conditionalize on $ch_{t+x}(A) < 1$. By SR, $Pr(A | ch_{t+x}(A) < 1) = 1$. This violates Principal, so we cannot have $Pr(ch_{t+x}(A) < 1) > 0$; thus $Pr(ch_{t+x}(A) < 1) = 0$, so $Pr(ch_{t+x}(A) = 1) = 1$, which is SC. Now assume SC. Then $Pr(A) = Pr(A | ch_{t+x}(A) = 1)$, because conditionalizing on a set of measure one does not change probabilities. By Principal, $Pr(A | ch_{t+x}(A) = 1) = 1$, so $Pr(A) = 1$, which is SR.

NOTES

* We are grateful to Nuel Belnap, Cristina Bicchieri, Susan Sterrett, Richmond Thomason, Michael Thompson, and two anonymous referees for useful discussion. For any errors that may remain, each of us blames the other guy.

¹ The word 'avowal' is often taken to suggest a performative of some sort; we use it in the absence of a better word to refer to self-ascriptions of attitudes.

² See especially van Fraassen (1985, §I; 1989, Chap. 7).

³ Dutch Books are often divided into two varieties: synchronic Dutch Books and diachronic Dutch Books. Synchronic Dutch Books involve a system of bets, all of which are purchased at the same time; the bets comprising a diachronic Dutch Book are purchased at different times. Van Fraassen's argument employs a diachronic Dutch Book. Note that a standard constraint placed on Dutch Book arguments is that the bookie be able to construct the book knowing only what the bettor knows. Thus a bookie who tricks a bettor into wagering on a horse race that has already occurred, and about whose outcome the bookie knows, may guarantee a loss for the bettor, but such a wager would not constitute a Dutch Book.

⁴ Invulnerability to *diachronic* Dutch Books is called dynamic coherence.

⁵ Here we ignore the possibility that such degrees of belief are not defined; see van Fraassen (1984, pp. 243–44) for discussion.

⁶ This means that $P_{a,t}$ has built into it what the agent a has learned (and remembered) up to and including time t . In particular, $P_{a,t}(A | p_{a,t+x}(A) = r)$ is *not* to be thought of as the agent's subjective probability for A , given that she knows *only* her future degree of belief.

⁷ Talbott (1991), Christensen (1991), and Maher (1992) offer similar examples. See below for further discussion.

⁸ Whether John is thereby committed to accepting such a wager if offered to him is another matter. The Dutch Book arguments seem to presuppose such a commitment; for a criticism of this commitment, see Bacchus et al. (1990).

⁹ See Teller (1973) for the argument.

¹⁰ See especially page 174.

¹¹ This maneuver is controversial, even among those who accept the standard Dutch Book arguments. The agent escapes a certain loss to the Bookie only by being so inscrutable that the bookie cannot predict the agent's deviations from conditionalization. But can unpredictability really be the salvation of the rational agent?

¹² Christensen (1991), Talbott (1991), and Maher (1992) maintain that diachronic Dutch Book arguments are unsound, while Bacchus et al. (1990) and Schick (1986) deny even the cogency of synchronic Dutch Book arguments.

¹³ Those uncomfortable with the notion of objective chance may find refuge in the subjectivist treatments of Skyrms (1980) or Lewis (1980). What is important for our purposes is not that ' $ch_{t+x}(A) = r$ ' refers to a single case propensity, but only to some objective facts that would render r the most appropriate probability for a to assign to A .

¹⁴ Lewis's version does not include explicit time indices, but the discussion in the postscript in Lewis (1986) makes it clear that the temporal indices do not distort Lewis's intent.

¹⁵ For more discussion about which propositions are admissible, the reader is referred to Lewis (1980).

¹⁶ The example is from van Fraassen (1984, p. 248).

¹⁷ For ease of notation we have omitted subscripts for the agent and the time.

¹⁸ The reader may check that Confidence' and Confidence'' are also violated.

¹⁹ The physical analogy is strict. An idealized teeter-totter consisting of point masses will balance if $\sum_t t \times (\text{Mass at } t \text{ units to the left of the fulcrum}) = \sum_t t \times (\text{Mass at } t \text{ units to the right of the fulcrum})$.

²⁰ This theme is developed more thoroughly in Daston (1988).

²¹ This has nothing to do with the Balance Principle of Section 2.

²² Although in the spirit of Lewis's ideal, Sobel's is more explicitly tailored to satisfy Reflection. Lewis's allusion to perception suggests an ideal designed for updating by conditionalization. Sobel (1990) articulates the latter ideal in greater detail.

²³ See van Fraassen (1984, pp. 250–52; 1990) for a discussion of probabilistic representation.

²⁴ Notice further that Reflection need not require of Rational Joe that he discover and correct every proclivity to over- or underestimate chances; it requires only that he not both: (i) discover himself to have such a proclivity, and (ii) fail to correct it. Rational Joe can adhere to Reflection by avoiding the first conjunct. Nor need we suppose that Rational Joe consciously adheres to Reflection – all we claim is that Reflection provides a true description of Rational Joe's belief states.

²⁵ If one has a conditional degree of belief of the form $P_t(A|p_{t+x}(A) = r) = s$, with $r \neq s$, then one is irreflective. Let us say that one who has such a belief, and then takes steps to realize the degree of belief on the right-hand side of the stroke, would be implementing a belief about which she is irreflective. For example, if Mary were deliberately to ingest LSQ, she would be implementing a belief about which she is irreflective. Maher (1992) offers a formal account of when it is irrational to implement such beliefs, which should prove useful in clarifying intuitions about this sort of example.

²⁶ This confusion is no doubt exacerbated by the Dutch Book arguments themselves, which seem to show that incoherence is irrational because it leaves one vulnerable to

financial loss at the hands of the Dutch bookie. But, of course, the actual prospects for suffering such a loss are slight indeed, for the incoherent agent can always abstain from gambling. If the Dutch Book arguments are cogent at all, they must be understood as illustrating, in a forceful way, an underlying epistemic defect in incoherent systems of belief.

²⁷ Armendt (1992) suggests a similar line of defense against the counterexamples.

²⁸ See note 25 above for the meaning of this expression.

²⁹ More precisely, the expected utility of implementing a belief about which one is reflective is *never less than* the expected utility of implementing a belief about which one is irreflective.

³⁰ Although most utterances of CPC would probably be interpreted as making an indirect speech act such as CP.

³¹ Van Fraassen seems to have something like this idea of minor commitment in mind in the following passage: "Having made a promise, I also have some obligation to prevent circumstances that would make it impossible to keep the promise. Having decided on a program of regular exercise, I have obliged myself to some extent to prevent travel arrangements, hangovers, lack of proper clothes and shoes, and so forth, that would interfere" (1984, p. 254). Van Fraassen does not develop the idea for conditional promises, however.

³² The conditional r -wager on A is one that is called off if the condition does not obtain, and becomes an ordinary r -wager on A otherwise. The conditional r -wager on A is equivalent to two unconditional wagers, on A & $p_{a,t+x}(A) = r$ and on $p_{a,t+x}(A) \neq r$.

³³ The argument is similar if $s < r$.

³⁴ See Searle and Vanderveken (1985), in which a speech act's degree of strength is one of the seven criteria according to which speech acts are taxonomized.

³⁵ We here make use of the distinction between the first-person forms of the copula verb: 'shall' and 'will'. According to grammar books, 'shall' is used to express a future indicative, whereas 'will' is used to express a commitment. This convention is falling into disuse in standard, spoken English.

³⁶ See Ginet (1979).

³⁷ This assumes that the distribution of Pr over propositions of the form $ch_{t+x}(A) = s$ has the form of a weighting function, putting non-zero weights on at most countably many propositions of this form. If the distribution has the form of a density function, the situation is somewhat more complex. Since, for any s , $Pr(ch_{t+x}(A) = s)$ will be zero, the conditional probabilities cannot be defined in the usual fashion. Let \mathcal{G} be a σ -field containing all the sets $\{ch_{t+x}(A) \in (a, b) : a, b \in [0, 1]\}$. Then, for fixed A , there exists a density function $Pr[A|\mathcal{G}]$, such that $\int_G Pr[A|\mathcal{G}] d(Pr) = Pr(A \cap G)$ for all $G \in \mathcal{G}$. Any two such functions differ at most on a set of measure zero. $Pr[A|\mathcal{G}]$ is a generalization of the standard conditional probability: if $G \in \mathcal{G}$ has no subsets in \mathcal{G} , and $Pr(G) \neq 0$, then $Pr[A|\mathcal{G}] = Pr(A|G)$ on G . Reflection then requires that $\int Pr[A|\mathcal{G}] d(Pr) = r$, where the integral is taken over the entire probability space. Thus, while $Pr[A|\mathcal{G}](\omega)$ takes on different values when ω belongs to different elements of \mathcal{G} , its values will be centered on r . This entails a version of balance in which the summations are replaced by integrals. This does not alter the informal understanding of Balance exploited in the text.

³⁸ This assumes that the proposition ' $p_{a,t+x}(A) = r$ ' is admissible in Lewis's sense. This should not be problematic, since it includes no information about times later than $t + x$.

It may be objected that it would beg the question to claim that ' $p_{a,t+x}(A) = r$ ' does no more than report a historical fact, because that assumes that assertions of subjective probability are autobiographical reports. But even if this proposition is understood as reporting the undertaking of an epistemic commitment, it should not be relevant to the outcome in a way that would prevent it from being screened off by the proposition reporting the objective chance. Note that, for $r \neq s$, the left-hand factor could never become one's credence function, for Reflection and Principal jointly rule out $P_{a,t}(ch_{t+x}(A) = s \ \& \ p_{a,t+x}(A) = r \neq s) = 1$ (see van Fraassen, 1984, pp. 246–48).

³⁹ Equivalently, this says that $E(ch_{t+x}(A)|p_{a,t+x}(A) = r) = r$, where $E(\cdot| -)$ is conditional expectation.

REFERENCES

- Armendt, B.: 1992, 'Dutch Strategies for Diachronic Rules: When Believers See the Sure Loss Coming', in D. Hull, M. Forbes, and K. Okruhlik (eds.), *PSA 1992*, Vol. I, Philosophy of Science Association, East Lansing, pp. 217–29.
- Bacchus, F., H. Kyburg, and M. Thalos: 1990, 'Against Conditionalization', *Synthese* 85, 475–506.
- Christensen, D.: 1991, 'Clever Bookies and Coherent Beliefs', *The Philosophical Review* 100, 229–47.
- Clifford, W. K.: 1879, *Lectures and Essays*, Macmillan, London.
- Daston, L.: 1988, *Classical Probability in the Enlightenment*, Princeton University Press, Princeton.
- Gigerenzer, G., Z. Swijtink, T. Porter, L. Daston, J. Beatty, and L. Krüger: 1989, *The Empire of Chance*, Cambridge University Press, Cambridge.
- Ginet, C.: 1979, 'Performativity', *Linguistics and Philosophy* 3, 245–65.
- Lewis, D. K.: 1980, 'A Subjectivist's Guide to Objective Chance', in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Vol. II, University of California Press, Berkeley, pp. 263–94 (reprinted with Postscripts in Lewis: 1986, *Philosophical Papers*, Vol. II, Oxford University Press, Oxford, pp. 83–132.)
- Maher, P.: 1992, 'Diachronic Rationality', *Philosophy of Science* 59, 120–41.
- Ross, W. D.: 1930, *The Right and the Good*, Clarendon Press, Oxford.
- Schick, F.: 1986, 'Dutch Bookies and Money Pumps', *Journal of Philosophy* 83, 112–19.
- Searle, J., and D. Vanderveken: 1985, *Foundations of Illocutionary Logic*, Cambridge University Press, Cambridge.
- Skyrms, B.: 1980, *Causal Necessity*, Yale University Press, New Haven.
- Sobel, J. H.: 1987, 'Self-doubts and Dutch Strategies', *Australasian Journal of Philosophy* 65, 56–81.
- Sobel, J. H.: 1990, 'Conditional Probabilities, Conditionalization, and Dutch Books', in A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990*, Vol. I, Philosophy of Science Association, East Lansing, pp. 503–15.
- Talbott, W. J.: 1991, 'Two Principles of Bayesian Epistemology', *Philosophical Studies* 62, 135–50.
- Teller, P.: 1973, 'Conditionalization and Observation', *Synthese* 26, 218–58.
- Van Fraassen, B. C.: 1984, 'Belief and the Will', *The Journal of Philosophy* 81, 235–56.
- Van Fraassen, B. C.: 1985, 'Empiricism in the Philosophy of Science', in P. M. Church-

land and C. A. Hooker (eds.), *Images of Science*, University of Chicago Press, Chicago, pp. 245–308.

Van Fraassen, B. C.: 1989, *Laws and Symmetry*, Clarendon Press, Oxford.

Van Fraassen, B. C.: 1990, 'Figures in a Probability Landscape', in J. M. Dunn and A. Gupta (eds.), *Truth or Consequences*, Kluwer, Dordrecht, pp. 345–56,

Green:

Department of Philosophy
University of Virginia
Charlottesville, VA 22903
U.S.A.

Hitchcock:

Department of Philosophy
Rice University
Houston, TX 77251
U.S.A.