8

Putting a Bridle on Irrationality: An Appraisal of van Fraassen's New Epistemology

Stathis Psillos

1. Introduction

Over the last twenty years, Bas van Fraassen has developed a 'new epistemology': an attempt to sail between Bayesianism and traditional epistemology. On his (1989) reading, Bayesian epistemology takes rationality to consist in rule-following, where the only rule of belief-revision is conditionalization. This is a logical (that is, non-ampliative) rule. It is meant to leave nothing (but the point of departure—that is, the prior probabilities) to our choice, but renders ampliative rules irrational. Traditional epistemology is more of a mosaic of views than one solid theory. On van Fraassen's (2000) reading, it too is committed to the view that rationality requires rule-following, but the rules include substantive ampliative ones (induction or inference to the best explanation). Traditional epistemology too is meant to leave nothing to our choice: belief and belief-revision require justification and this is effected by substantive principles of rationality and (ampliative) rules.

Van Fraassen is dissatisfied with both approaches. He calls his own alternative 'voluntarism'. For him, it is rational to form beliefs that go beyond the evidence, but these beliefs are not rationally *compelling* by virtue of substantive principles

and ampliative rules. 'Belief' van Fraassen says, 'is a matter of the will' (1984: 256). It involves decision, cognitive commitment, intention, and engagement. A constant pillar of his voluntarism is the thought that rationality involves permission rather than obligation. As he (2002: 101) notes, his conception 'of what is rational or rationally endorsable. . .is entirely at odds with the traditional "compelled by reason" conception'.

The present chapter aims to offer an appraisal of van Fraassen's conception of rationality. It must be noted that his views on rationality are quite independent from his views on constructive empiricism. In fact, one can be a scientific realist and adopt van Fraassen's conception of rationality: belief in electrons and so on may well come out as rational under van Fraassen's conception of rationality. But so may disbelief in them (or, agnosticism about them). Hence, van Fraassen's conception of rationality is *suitable* for constructive empiricists in that it shows that belief solely in the empirical adequacy of theories *is* rational (cf. 2001: 162, 168). Besides, van Fraassen's conception of rationality makes constructive empiricism *safe*: belief in scientific realism is *not* rationally compelled. I shall not concern myself with the issue of scientific realism. I will focus my attention on the general implications of van Fraassen's views for the concept of rationality.

The structure of the chapter is this. In Section 2, I review the Bayesian *structural* conception of rationality and argue that it has been found wanting. In Sections 3 and 4, I analyse van Fraassen's voluntarism. I raise some objections about van Fraassen's reliance on prior opinion and argue that the *content* of a belief matters to its rationality. In Section 5, I criticize van Fraassen's view that inference to the best explanation is incoherent. Finally, in Section 6, I take on van Fraassen's conception of rationality and show that it is too *thin* to capture rational judgement fully.

2. Structural Rationality

There are two ways to view Bayesianism: I'll call them *synchronic Bayesianism*, and *diachronic Bayesianism*.

Synchronic Bayesianism takes the view that the axioms of the probability calculus are an extension of ordinary deductive logic. The demand for probabilistic coherence among one's degrees of belief is a *logical* demand: a demand for logical *consistency*. On this view, defended by Howson (2000), the

136 / **Putting a Bridle on Irrationality**

degrees of beliefs that an agent possesses should, at any given time and on pain of inconsistency, satisfy the axioms of the probability calculus. Otherwise, she is subject to a (synchronic) Dutch book, that is, to a set of synchronic bets such that they are all fair by her own lights, and yet, taken together, make her suffer a net loss.[1] Howson takes this kind of logicized Bayesianism to be fully dissociated from a theory of rationality and of rational belief. For him, logic is about consistency and 'not *about* rational belief or action as such' (2000: 133). The view that synchronic probabilistic coherence is a canon of rationality cannot be maintained, according to Howson, since it would require a non-question-begging *demonstration* that any violation of the axioms of the probability calculus is positively irrational. But no such proof is forthcoming.

What is remarkable about synchronic Bayesianism is that no pretension is made for offering a logical recipe for belief-revision (or, better, for degree-of-belief revision). In particular, there is no logical requirement for belief-updating by means of conditionalization on the evidence. Howson is adamant that if people update their degrees of belief by non-conditionalizing on the evidence, they don't thereby violate any canon of rationality.

Diachronic Bayesianism places conditionalization (either strict, where the probability of the learned evidence is unity, or Jeffrey, where the evidence one updates on can have probability less than 1) on centre stage. It is supposed to be a canon of rationality (certainly a necessary condition for it) that agents should *update* their degrees of belief by conditionalizing on evidence: $\text{Prob}_{\text{new}}(-) = \text{Prob}_{\text{old}}(-/e)$, where e is the total evidence. The penalty for not doing this is liability to a Dutch book *strategy*: the agent can be offered a set of bets *over time* such that a) each of them taken individually will seem fair to her at the time when it is offered; but b) taken collectively, they lead her to suffer a net loss, come what may.[2] As is now generally recognized, the penalty is there on a certain condition, namely, that the agent *announces in advance* the method by which she changes her degrees of belief, when new evidence rolls in, *and* that this method is different from conditionalization (cf. Earman 1992: 47). Critics of diachronic Bayesianism (which include some advocates of synchronic Bayesianism, for example, Howson) are quick to point

---

[1] The monetary aspect of the standard renditions of the Dutch book theorem is just a dramatic device. The thrust of the Dutch book theorem is that there is a structural incoherence in a system of degrees of belief which violate the axioms of the probability calculus. For more on this, see Skyrms (1984: 20−3) and Lange (1999).

[2] This is the famous Lewis−Teller argument (see Teller 1973).

out that there is no general proof of the conditionalization rule (cf. Earman 1992: 46−51). In fact, as Howson (2000: 136) notes, there are circumstances under which conditionalization is an *inconsistent* strategy. When an agent is in a situation in which she contemplates her $Prob_{new}(-)$, she is in a *new* and different (betting) situation in which the previous constraints of $Prob_{old}$ need not apply. A case like this is when the learning of the evidence *e* does upset the conditional probability $Prob(-/e)$. Indeed, when the learning of *e* does not cause any changes in the agent's $Prob(-/e)$, then conditionalization is mandatory (cf. Howson 2000: 139). Diachronic Bayesianism has a point. Under *certain* circumstances, an agent should update her degrees of belief by conditionalizing on the evidence. But it does *not* follow from this that Bayesian updating is a canon of rationality.[3]

In so far as diachronic Bayesianism succeeds as a theory of rationality (and as we have just noted this is by no means obvious), it offers a *structural conception of rationality*: rationality pertains to the structure of a belief system and not to its content. This conception is actually shared by synchronic Bayesianism too. The difference is that synchronic Bayesianism looks at the belief-structure *at* a time and not *over* time. It does not matter *what* you believe (that is, what the propositional content of your beliefs is—provided that it's not contradictory). All that matters is how what you believe hangs together (at a certain time, or over time). According to the Bayesian structural conception of rationality, it is *not* irrational to maintain unjustified opinion. It is well known that for subjective Bayesians prior opinion can come from anywhere. And so can the prior probabilities. This is a natural consequence of the thought that rationality does not pertain to the *content* of the opinion or belief.[4] The standard (subjective) diachronic Bayesian picture is that people start with some prior opinion (as a 'free move' to which they are *entitled* without justifying it (Lange 1999: 303)) and then update it by conditionalizing on the evidence. This is purely *logical* updating. It's not ampliative. It does not introduce new content; nor does it modify the old one. It just assigns a new probability to the old opinion.[5]

[3] For a sustained critique of the view that rationality requires diachronic conditionalization, see David Christensen (1991).

[4] I am not suggesting that structural constraints never place restrictions on the content of a belief. Certain propositions must be believed or not be believed because, ultimately, they have a certain structure (for example, *p* or not-*p*, and *p* & not-*p*, respectively). But structural constraints will never constrain the content of an atomic proposition and of many a molecular proposition.

[5] It is interesting to note that a version of the structural conception of rationality was also advocated by Karl Popper and his fellow critical rationalists.

138 / **Putting a Bridle on Irrationality**

This aspect of Bayesianism might bring to light its greatest shortcoming as a purported theory of rationality: its radical incompleteness.[6] Without supplementation, Bayesianism neglects the role of *evidence* in rational belief. This might sound paradoxical, given that diachronic Bayesianism is meant to be a theory of belief-updating, *given* the evidence. But it is not. Diachronic Bayesianism dictates how probabilities should be redistributed over the elements of a belief-corpus, *if* and *when* a new belief (in this case, a belief about what the evidence is) is about to become part of the belief-corpus. But it says nothing about when a new belief *should* be accepted and become part of a belief-corpus.[7]

Bayesians might think that this is just fine, given that they dissociate the rationality of a belief from its content. They may reply that, according to the Bayesian theory of confirmation, the fact that a certain proposition e is evidence for a hypothesis H is fully captured by the following relation: $\text{prob}(H/e) > \text{prob}(H)$. This, however, is an incomplete response for the following reason. When we think about evidence, there are two things that we need to think about. The first is what the relation —*is evidence for*— consists in. Bayesian confirmation does address this issue. The second thing is the nature of the first *relatum* of the above relation (cf. Williamson 2000: 189). Here the task is not just to investigate what kinds of things can be evidence (i.e., whether they are propositions, whether they relate to observations etc.). The task is also to look into the epistemic status of whatever *is evidence for*—. It is this issue that Bayesianism fails to address.[8] I am persuaded by Williamson (2000: 194–200)

---

[6] I leave aside another problem that Bayesianism faces, viz., 'the problem of incorrigibility' (Lange 1999: 300). Bayesian agents seem enslaved to their prior probability distribution: degree of belief updating merely alters the probability agents assign to the propositions that express their initial beliefs. This feature of Bayesianism seems to conflict with the fact that agents radically revise or abandon some of their prior beliefs. As Brad Monton reminded me, it is possible that an agent believes those propositions to which she assigns high probability. Conditionalization on fresh evidence may then result in lowering her probability for a proposition, leading her from believing that P to not believing that P. However, this kind of move would be at odds with strict Bayesianism, since it implies that there is a cut-off point in which the evidence renders a proposition belief-worthy or disbelief-worthy.

[7] This might well be a variant of a well-known objection to Bayesianism, viz., that its reliance on purely subjective prior probabilities fails to capture the all-important notion of rational or reasonable degrees of belief. For a critique along these lines, see Earman (1992: 57–9).

[8] There is a notable exception. Teller (1973: see especially 238–40) relies on a *causal* account of observation in order to show how conditionalization on the evidence and a substantive account of what constitutes evidence can be put together.

that all evidence is propositional. But be that as it may, Bayesians remain silent on when it is rational to accept something as evidence and when it is rational to take pains to accommodate the evidence within one's belief-corpus. For instance, it is entirely open to Bayesians to argue that some (perhaps all?) evidence can be neglected. But this cannot be generally right. Though I shall discuss this issue in some detail in Section 6, it is pertinent to say the following. There is a lot of evidence *that the earth is (roughly) round* and *very* little (if any) evidence *that the earth is flat*. Yet, one could be a perfectly consistent Bayesian agent, even if one believed that the earth is flat. There seems to be nothing in Bayesianism which would render irrational an agent who neglected evidence that points to the roundness of earth in order to safeguard her belief that the earth is flat. In fact, a Bayesian agent could rationalize her attitude by giving zero prior probability to the hypothesis that the earth is round.[9]

3. Enter van Fraassen

In trying to lay out as clearly as possible his difference from Bayesianism, scepticism, and traditional epistemology, van Fraassen (1989: 178) states the following four basic epistemic principles:

(I) There can be no independent justification to continue to believe what we already find ourselves believing.

(II) It is irrational to maintain unjustified opinion.

(III) There can be no independent justification for any ampliative extrapolation of the evidence plus previous opinion to the future.

(IV) It is irrational to extrapolate ampliatively without justification.

Endorsement of all four positions amounts to scepticism, he says. Orthodox Bayesianism accepts I, III and IV. It avoids scepticism by denying II. The Bayesian's insistence that there is no substantive theory of rational-objective prior degrees of beliefs, 'allows him [the Bayesian] to live a happy and useful life by conscientiously updating the opinions gained at his mother's knees, in response to his own experience thereafter' (1989: 178). Van Fraassen's own view is neither sceptical nor Bayesian because he endorses I and III but rejects II

---

[9] The need to go beyond a purely structural conception of rationality is highlighted by Worrall (1993) in relation to attempted Bayesian solutions to the Duhem−Quine problem.

140 / **Putting a Bridle on Irrationality**

and IV. Unlike the Bayesians, van Fraassen *denies* that it is necessarily irrational to be involved in ampliative extrapolation from the evidence. This mixture of theses is the kernel of van Fraassen's 'new epistemology'. More specifically, 'new epistemology' has it that Bayesian conditionalization—a non-ampliative rule—is *not* a rationally compelling way to update one's previous opinion, given the evidence. Van Fraassen (1989: 175) is clear on this:

Like the Bayesian I hold that rational persons with the same evidence can still disagree in their opinion generally; but I do not accept the Bayesian recipes for opinion change as rationally compelling.

Unlike the Bayesians, van Fraassen (1989: 174) thinks that

rationality does not require conditionalization, nor does it require any commitment to follow a rule devised beforehand.

Van Fraassen calls his new epistemology 'voluntarist'. Traditionally, voluntarism is the view that having a belief is something that a person does voluntarily and can control. But it is also equated with the kindred view that there can be reasons to believe that are *not* evidential: one can come to believe that *p* (i.e. one can decide to believe that *p*) on the basis of reasons that are not related to the probability of *p* being true (or, equivalently, on the basis of reasons that do not enhance its probability of being true).

There is a rather decisive argument against voluntarism. According to Bernard Williams (1973), it is (pragmatically) incoherent to say that I believe at will. Belief aims (constitutively) at truth. If I could acquire a belief at will, then I could acquire it whether it was true or not. Being *my* belief, I take it to be *true*. But I also *know* that my belief could be acquired whether it was true or not. Hence, I am (pragmatically) incoherent. I am saying: I believe that *p* (is true) but I believe that *p* whether it is true or not.[10] The second conjunct is in conflict with the first, since it severs the link between belief and truth. Put in a different way, Williams's claim is that the state I am in when I acquire a belief at will is *not* a belief—for belief is normatively connected to truth and to the evidence that supports it. Of course, I might consciously follow a Pascal-wager type of strategy to cultivate a certain belief. But as Williams (2002: 83) has

FN:10

---

[10] This is not a formal contradiction, as it can be easily seen if we replace the 'I' with a 'she': 'she believes that *p* (is true) but she believes that *p* whether it is true or not' might well be true. Yet, when this sentence is uttered by *me*, it is (pragmatically) incoherent.

recently noted, a certain requirement for this strategy is that 'I must be able to forget that this is how I acquired the belief, or if I remember that I acquired it in this way, I need an explanation of how that is supposed to be connected with the belief's being true.' The point is clear: so called non-epistemic reasons for belief had better be disguised or internalized as *epistemic reasons*, that is, as reasons that have to do with the truth of the belief (see also Foley 1993: 17−18).

In light of this, what is the shape of van Fraassen's voluntarism? Van Fraassen does *not* say that we can believe just any proposition at will (1984: 236 note 3). Nor does he say that we can coherently assert that we believe a proposition and that we believe it for reasons that do not make it more likely to be true (cf. 2002: 89; 2001: 167).[11] Consequently, his voluntarism is not directly threatened by Williams's argument. In fact, there is a sense in which van Fraassen *must* think that *crude* voluntarism is false. Do I now have an option *not* to believe that I am reading this chapter? Some beliefs are certainly forced on us. Consider one well-known passage of his: 'we can and do see the truth about many things: ourselves, others, trees and animals, clouds and rivers—in the immediacy of experience' (1989: 178). I think the best (only?) way to interpret this is that some truths are indeed forced on us—we cannot choose not to believe in them. Consequently, van Fraassen's 'voluntarism' (a term which, as he says, he uses with 'minimal connotations') is a rather subtle position. It does consist in an attempt to give 'central importance to the will and the role of decision' (2002: 77), but it does not entail that agents can believe anything they want. A useful way to think of van Fraassen's voluntarism may be in terms of what he calls 'an epistemic policy': 'If we choose an epistemic policy to govern under what conditions, and how far, we will go beyond the evidence in our beliefs, we will be setting down certain boundaries' (1985: 254). Epistemic policies are not dictated by the evidence, van Fraassen thinks, and they involve certain decisions and commitments (for example, where to set the boundaries of experience or where to stop seeking further evidence or where to start withholding belief). But let us look more carefully at the elements that comprise van Fraassen's voluntarism.

[11] Though van Fraassen draws a distinction between belief and acceptance (which is more relevant when it comes to one's attitudes towards theories than to one's opinions on 'little matters of fact' (2002: 90)), he resists the idea that acceptance differs from belief in that the former, but not the latter, can be held for non-epistemic reasons (cf. 2001: 166−7; 2002: 89−90).

4. Voluntarism

Van Fraassen's voluntarism rests on two theses. The *first* concerns 'the status of judgement' (1989: 179). A judgement is not an 'autobiographical statement of fact'. As he (1989: 179) puts it:

[A judgement] does not state or describe, but avow: it expresses a propositional attitude. To make it is to take a stand. To adopt an attitude or a stance is akin to commitment, intention.

This is a crucial move. Van Fraassen points to a difference between the first-person perspective and the third-person one. Compare the following two statements:

(A) *X*'s opinion is that we humans have descended from apes.
(B) It is *my* opinion that we humans have descended from apes.

The person *X* in (A) might be myself. Yet, in a number of places (1984: 253; 1989: 179; 1995a), van Fraassen argues that there is a *crucial* difference between (A) and (B). (A) asserts what *X* believes. (B) may be taken to assert exactly the same. But, appearances to the contrary, (B) asserts more than (A). When (A) is asserted by *me*, it is not an autobiographical statement. (B) expresses *my* opinion, and with it, it implies a certain commitment or intention on my part. It involves a 'decision' on my part to commit myself to a certain stance or to follow a certain course of action (cf. 1984: 254). Van Fraassen uses this view to defend his Reflection Principle.[12] But this need not concern us here. What is relevant is that van Fraassen takes epistemic judgements not to be purely factual.

The *second* thesis of van Fraassen's voluntarism concerns the concept of rationality. Van Fraassen introduces his views of rationality by drawing a distinction between the 'Prussian' concept of rationality and the 'English' one. According to the first, 'everything is forbidden which is not explicitly permitted', while according to the second 'everything is permitted that is not

---

[12] This principle says that 'the agent's present subjective probability for a proposition A, on the supposition that his subjective probability for this proposition will equal *r* at some later time, must equal this same number *r*' (van Fraassen 1984: 244). This principle has been defended by van Fraassen in his (1995a) and has been criticized by Christensen (1991), Plantinga (1993), and Howson (2000).

explicitly forbidden' (1989: 171). Van Fraassen opts for the English conception. Accordingly (1989: 172−3),

what it is rational to believe includes anything that one is not rationally compelled to disbelieve. And. . .the rational ways to change your opinion include any that remain within the bounds of rationality.

Let's call this van Fraassen's *central dictum*. In a slogan: 'Rationality is only bridled irrationality' (ibid., cf. also 1983: 299).

This view has been one of the constant pillars of his thought (cf. 1985: 248; 2002: 92, 97) It is voluntarist because, as he explains, it leaves an irreducible element of free choice (1989: 176). Since what is not required is not always forbidden, it is, ultimately, up the agent's *free choice* to decide what to believe and how to change her beliefs. Rationality, ultimately, is a concept of permission and not of obligation. Still, rationality is bounded. As van Fraassen says, '[belief changes] are rational exactly if they are rationally permitted, if they do not transgress the bounds of reason' (2002: 92).

So reason has *bounds* (cf. also 1985: 248). Elsewhere, van Fraassen talks of 'the dictates or criteria of reason' (2002: 97). We shall discuss these bounds in some detail in Section 6. For now, it is important to stress that for van Fraassen rationality is bounded by two items (so to speak). One is *prior opinion*. The other is *logical and (synchronic) probabilistic consistency* among one's beliefs. There seems to be a third item. This is what one might call the *no-self-sabotage principle* (cf. 1985: 248; 1989: 157, 347). As van Fraassen (1989: 157) puts it:

a minimal criterion for reasonableness is that *you should not sabotage your possibilities of vindication beforehand.*

Elsewhere, he says: 'A decision is unreasonable if vindication is *a priori* precluded' (1983: 297). An action, opinion, decision, van Fraassen says, may be reasonable without having been vindicated. Whether it is vindicated or not depends on its outcome, and hence, broadly speaking, on the way the world turns out to be. Still, it would be unreasonable for an agent to put herself in a situation in which vindication is a priori impossible. In a way, the *no-self-sabotage principle* is an operationalization of the requirement of consistency. A logically inconsistent belief-corpus can *never* be vindicated. Similarly, a synchronically incoherent belief-corpus sabotages itself (remember the Dutch Book argument). But the *no-self-sabotage principle* relates to belief-change, too. *If* one is to follow a rule for belief change, one would sabotage oneself if this rule was different

144 / **Putting a Bridle on Irrationality**

from Bayesian conditionalization (cf. 1989: 173—4, 322, 347). Yet, the *if* above is really important for van Fraassen. For one is *not* rationally compelled to follow a rule in belief change (cf. 1989: 347). There is another case of self-sabotage that needs to be noted. Calibration, van Fraassen (1983: 300) says, is 'a measure of how reliable one's judgements have been as an indicator of actual frequencies'. A weather forecaster, for instance, is well-calibrated if when his judgement is that the chance of rain today is 0.8, it turns out that 80 per cent of the days like today (in relevant respects) were rainy. Take, now, vindication to be calibration, that is, a judgement is vindicated (successful) if it is perfectly calibrated. Then van Fraassen (1983) shows that an agent's degrees of belief are perfectly calibrated only if they satisfy the axioms of the probability calculus. An agent whose degrees of belief are incoherent cannot be perfectly calibrated. This is a far-reaching result.[13] It gives a reason to strive for coherence. But it can also accommodate a sense in which an opinion can be *right*. We shall come back to this issue and its implications in a moment.[14]

Note that van Fraassen's own conception of rationality is to *some extent* structural. As his *central dictum* implies, what is rationally permitted to believe is constrained by what is *rationally forbidden* to believe or disbelieve. But if there are things that one is rationally forbidden to believe or disbelieve, there must be at least some rules that issue in these prohibitions. The rules themselves are *structural*: they concern only the coherence of a body of beliefs. Since they concern *only* how beliefs hang together at one time (or perhaps, over time), they will make the *content* of one's beliefs irrelevant to ascriptions of rational belief. Van Fraassen (2001: 168) is clear on this:

Coherence means: no self-sabotage. The constraints of coherence are really empty, because they don't limit factual content of belief at all. The yoke is easy. The burden is light.

---

[13] I have over-simplified. Van Fraassen explicates vindication in terms of *potential* calibration. This notion relates to the fact that we have to take into account limiting relative frequencies. But all this need not concern us here. The interested reader should look at Lange (1999).

[14] Lange has offered a new reading of Bayesian conditionalization, according to which conditionalization governs 'the steps in the *arguments* by which our current opinions are to be justified by the evidence that we have *already* assembled' (1999: 295). He shows that if justificatory arguments violate Bayesian conditionalization, they fail to be calibrated. This is a big step forward in the debate since it shows how the evidence *justifies* an agent's current opinion.

Yet, according to van Fraassen, the subject's beliefs (that is, the content of her belief) will be constrained by the subject's *prior opinion*.[15] This is contentful. Hence his conception of rationality is *not* purely structural.

4.1 Prior Opinion and Calibration

It's important at this point to look into the role of *prior opinion* in van Fraassen's conception of rationality. In line with the first thesis of his voluntarism, prior opinion is not a body of autobiographical beliefs. Rather, it expresses a certain set of commitments and an intention to stand by them. It's a set of commitments to view the world in a certain way. There is an obvious worry about all this, which van Fraassen (1989: 175, 178−80) takes pains to disarm: doesn't he succumb to relativism? A relativist would argue as follows: since no body of prior opinion is rationally compelling, all prior opinions are equally (un)justified; hence, anything goes. As we have seen, van Fraassen *denies* that prior opinion requires justification. He also denies that if there is no logical compulsion, there is no justification. So he denies the premises of the relativist argument. In any case, the following thing is clear. Van Fraassen denies truth-relativism. He insists that prior opinion is constrained, in a non-trivial way, by an *external* constraint, namely, truth. There is, he (1989: 177) says, an 'objective criterion of rightness' of opinion:

Certainly our opinion is right or wrong, and this depends on what the world (the facts we make judgements about) is like. (1989: 177)

The truth of a judgement (for example, about unobservable entities) might not be decidable. Yet, a certain judgement may be true 'if only by accident' (1989: 177).

This 'objective criterion of rightness' can show how there can be traffic between an autobiographical statement of fact and the commitment to an opinion. Even if truth is not always decidable, there are many occasions (in science and in everyday life) that truth and falsity *can* be decided. In those cases (which, as van Fraassen (2000: 273) rightly notes, require some luck), a disagreement can be purely factual. More to the point, in such cases, prior opinion can be vindicated or corrected. It's no longer optional for someone to

---

[15] There are exceptions and van Fraassen discusses them thoroughly in his (2002). They concern revolutionary episodes, when a new theory or hypothesis is envisaged. Then, the agent should (and does) revise her prior opinion.

146 / **Putting a Bridle on Irrationality**

say: this is *my* opinion and I will stand by it no matter what. This non-trivial criterion of correctness shows that van Fraassen's distinction between an autobiographical statement of fact and the commitment to an opinion is not so watertight. For, I can certainly move (in reflective mode) between a first-person perspective on my opinion and a third-person one. I can certainly ask of *my* opinion 'does it resonate with the facts?' If what the (relevant) facts are can be known (as it certainly can on very many occasions), then I can use this *knowledge* to answer the foregoing question. The point is not that people can be criticized for having false beliefs. Rather, the point is that it is irrational for people to entertain some beliefs, if they can do a little more work to find out whether their beliefs are false.[16]

These considerations can be strengthened if we take account of the need for our beliefs to be calibrated. The notion of calibration, you may recall, captures an important sense in which a belief (or opinion) is *right*. In fact, van Fraassen (1983: 301) notes that 'calibration plays the conceptual role that truth, or empirical adequacy, plays in other contexts of discussion.' Now, if a belief is (perfectly) calibrated, then it is not *just* right (that is, correct). In a sense, it is also the *rational* belief to have. Its rationality comes from the fact that, being perfectly calibrated, it maximizes the chances of vindication. To say of a belief that it is (perfectly) calibrated is to say of it that it is accurate. But accurate beliefs matter for the choice of effective strategies for achieving certain goals. If I am planning a long trip tomorrow, and if I don't feel comfortable when I drive for long with my daughter on board in the presence of heavy rain, then it matters for what I will do (and to what it is rational for me to do) to have an accurate belief about the chances of heavy rain. It will make a difference in my choosing to drive to my destination or to go there by train. Given my goals and other constraints, my opinion had better be calibrated. Acquiring accurate beliefs is the rational thing for an agent to do. Conversely, refraining from acquiring accurate beliefs is irrational for an agent, if she wants to follow effective strategies to achieve her goals. For, knowingly, she minimizes the chance for vindication. Now, an agent might not be able to get in advance information concerning the accuracy of her beliefs. Worse, there may be cases in which there is no such information available. But what about the host of

---

[16]   I have another worry about van Fraassen's reliance on prior opinion. It's not clear what sense to make of it. What *is* our prior opinion? What does it include? How is it to be circumscribed? And who are the *we* who have this prior opinion?

cases in which there is such information? It does not seem rational on the agent's part to ignore it. There are two morals I want to draw from this. The first is that the demand for calibration is a means to traffic between an autobiographical statement of fact and the commitment to an opinion. The second is that the demand for calibration shows how the *content* of an opinion matters to its rationality. Beliefs with certain contents cannot be perfectly calibrated, while beliefs with other contents can be.[17]

4.2 Evaluativism

At one point, van Fraassen claims: 'We believe that our beliefs are true, and our opinions reliable. We would be irrational if we did not normally have this attitude toward our opinion' (1989: 171). How much pause should this give us? Rationality requires reflection. It also requires epistemic responsibility. Finally, it requires responsiveness to reasons, as Nozick (1993: 71) put it. All these suggest that though prior opinion should be taken seriously, it can (and should) be subjected to criticism; not just externally, but also internally (that is, by the very subject who holds this opinion). If we are reflective about the content of our beliefs, then we need to consider what evidence supports them and not just how they cohere with the rest of what we believe. If we are epistemically responsible, we also need to consider the evidence that supports our beliefs: we should consider how we should conduct inquiry in a way that secures that our beliefs bear the weight of the evidence. If we want our beliefs to be responsive to reasons, then again we should examine how the relevant evidence supports the beliefs we have. Prior opinion does constrain our current beliefs, but the real issue, as van Fraassen (1989: 180) himself implicitly acknowledges, is whether it *should*. In trying to deal with this issue, it seems obvious that we have to look into the *content* of prior opinion. It's not enough to appeal to an external constraint (truth). Nor is it enough to adopt a stance which commits us to our prior opinion. What is required is subjecting prior opinion (ours and others') to critical scrutiny. Indeed, I think

---

[17] I am leaving aside an important issue, viz., how the reference class that determines the probability of a belief is to be specified. This issue matters for the notion of calibration. But I am willing to go along with van Fraassen's (1983) view that though the specification of the reference classes is a subjective matter, once a specification has been made, it is an objective matter whether or not one's degrees of belief are calibrated with relative frequencies. See also Foley (1993: 157). For a somewhat different take on the issue of reference classes, see Lange (1999).

148 / **Putting a Bridle on Irrationality**

that subjecting our prior opinion to critical scrutiny is a mark of rationality.[18] It might be relevant here to highlight a distinction that Nozick (1993: 70) draws between the act of believing that *p*, and the rationality of believing that *p*. The former might be constrained by prior opinion only. But the latter should be sensitive to relations of evidential support and responsive to reasons.

To forestall a possible misunderstanding, I am not saying that van Fraassen denies all this. In fact, in his (1989: 175) he stresses:

I have not implied that standards of criticism do not exist, but only that they are not a matter of logic.

In a sense, we ought to have known that all along, had we paid more attention to what Duhem taught us back in 1906.[19] The fact that standards of criticism are not a matter of *logic* does not imply that they are not rule-bound. Nor does it imply that they cannot be articulated and followed. My point really is not that van Fraassen denies that responsiveness to reasons and evidence should characterize a rational agent. Rather, it is that I don't see how exactly responsiveness to reasons and evidence is accommodated within his voluntarism. The bottom line is this: one should distinguish between the question of how reasons and evidence *can* be taken into account (for example, within the framework of Bayesian conditionalization) and the further question of whether and how exactly they *should* be taken into account.

In his (1989) as well as in his (2002), van Fraassen aligns himself with William James's idea that in forming our opinion we pursue two main aims: 'to believe truth and to avoid error' (1989: 172; 2002: 86). If our only aim was to believe truths, then we could believe everything, thereby making sure that all truths are caught within the net of our beliefs. If, one the other hand,

---

[18] This is one of the central messages of Foley's (1993) account of egocentric rationality.

[19] To cut a long story short, Duhem argued that there was space for rational judgements in science which was *not* captured by his official slogan, viz., *scientific method = experience + logic*. This space includes explanatory considerations, that is considerations which the official Duhemian dogma classified under the banner of metaphysics. As is well-known, Duhem recognized what Poincaré made famous by saying that though experience does not, strictly speaking, contradict a theory, it can *condemn* it. He also made the point that experience and logic cannot dictate how to revise theories in the face of recalcitrant evidence. If crucial experiments are 'impossible in physics' (1906: 188), then how do theories get abandoned? Any answer would have to go beyond the strict limits of experience and logic. And Duhem's own answer did. He appealed to other (explanatory) criteria of assessment. What he saw clearly was that the employment of such criteria was a) indispensable, and b) not algorithmic. Their exercise requires the employment and exercise of *judgement*.

our only aim was to avoid error, then we could make sure that we believe only tautologies. Searching after truth is a demand for informative beliefs. Avoiding error is a demand for secure beliefs. The two aims pull in contrary directions. And if both are valued, as they should be, there must be a balance between them. But the very notion of balance between two conflicting aims introduces an element of *value*. For, as van Fraassen (2002: 87) notes, this measure of balance is *not* an objective matter (even though achieving truth and avoiding error are): it is 'contextually qualified by our interests and values' (2002: 90). Once more, voluntarism comes in as a solution. For, this measure of balance depends, ultimately, on a value judgement (What do we value most and why? What truths are we interested in? What weights shall we attach to searching after truth and to error-avoidance?). As such, it is the outcome of the exercise of our *free* choice (cf. 2002: 87–8). The important point then is this. Van Fraassen's new conception of rationality has values and evaluation occupy centre stage. In fact, his reliance on value judgements ties together the two elements of his voluntarism. For as he says, value judgements do not 'simply make. . .autobiographical statements of fact, but. . .affirm or express the evaluative propositional attitude, or the commitment to those values' (1993: 23).

In an earlier draft of this chapter, I argued that van Fraassen's view is a species of what Field (2001) has called 'evaluativism'. Owing to lack of space, I won't pursue this issue here (but see van Fraassen 1993: 21–8). For convenience, I will just use the term 'evaluativism' to refer to van Fraassen's value-based approach to epistemology.[20] But I do want to point out that this value-based approach is consistent with the existence of some objective judgements of rationality. We have already seen one reason why: evaluativism is consistent with there being objective criteria of rightness. In fact, as we have seen, thanks to the 'objective criterion of rightness' some opinions may be irrational to entertain.

Another reason why evaluativism is consistent with there being objective criteria of rightness concerns the possibility of comparative judgements of rationality. Even if goals are not objective, how one might go about achieving one's goals may well be an objective matter. This is, of course, well-charted

---

[20] Van Fraassen's recent call to abandon the search for 'objectifying epistemology' (2002: 82) might be seen as a way to advance evaluativism. 'Objectifying' are the epistemologies that consist in a factual theory-writing project about cognitive functioning (2002: 76). One central element of van Fraassen's critique of this project is that it leaves out the role of values and of evaluation in characterizing the basic epistemic concepts and in explaining their role.

FN:21

territory. It relates to the well-known theme of normative naturalism and to the instrumental conception of rationality.[21] I am not going to expand on this issue here (see my 1999: 176−82). But the thrust is this: relative to a certain goal X, it is a factual issue whether method M is a reliable means to achieve X, or whether method M is more reliable than M' for X. There is, then, a factual way to make comparisons: some methods are better than others because they lead to more truth and less falsehood. Here again, the comparison presupposes some goals (or values). But this is what we should take to heart if we endorse evaluativism. Now, van Fraassen is a critic of instrumental rationality (1993: 22−3). I think he's right in so far as the advocates of instrumental rationality intend to eliminate the role of values. But it is an open question whether they do indeed intend to eliminate values.

5. Rule-Following

We have already seen that van Fraassen denies that rationality requires rule-following (1989: 322, 347). The relevant argument proceeds in two steps. The *first* step (1989: 160−70) aims to show that if one were to follow an ampliative rule for belief updating, (hence a rule which is not Bayesian conditionalization), one would be incoherent. The *second* step (1989: 174) aims to show that even Bayesian conditionalization is not required for rationality. This two-step argument motivates his new conception of rationality. The second step, in particular, saves the bacon of ampliative extrapolation: it is rational. Yet, on van Fraassen's views, it is *not* subject to rules that one is rationally compelled to follow and that leave nothing to one's choice.

There has been a lot of discussion of the first step in van Fraassen's argument. As is well-known, van Fraassen directs it against inference to the best explanation (IBE). The crux of IBE, no matter how it is formulated, is that explanatory considerations should inform (perhaps, determine) what it is reasonable to believe. Here is, in broad outline, how he attempts to show that following IBE as a rule is incoherent. Before Peter learns the evidence e, he uses standard Bayesian techniques to anticipate the posterior probabilities of several competing hypotheses, given e, including some hypothesis H. Then Peter learns that H is the best explanation of the evidence e. Peter is a friend

---

[21] For a thorough defence of instrumental rationality, see Foley (1993).

of IBE. That is, he thinks that the best explanation of the evidence should be favoured over others. In line with this, he decides to give a *bonus* to the posterior probability prob(H/e) of H, given that H is accepted as the best explanation of the evidence *e*. *After* having anticipated prob(H/e) in a Bayesian way, Peter sees this hypothesis as the best explanation of the evidence, and *raises* its posterior probability. Van Fraassen then shows that Peter is subject to a Dutch book strategy: he is subject to a set of bets (over time), which guarantee him a net loss, come what may. Peter is irrational because he violates the *no-self sabotage principle*: he commits himself to a rule for belief-change which makes him (diachronically) incoherent.

What exactly does van Fraassen's argument show? It certainly shows this: if one were to think of IBE as a rule which is parasitic on Bayesian conditionalization, in the precise sense of giving bonus probabilities to the hypothesis that is accepted as the best explanation of the evidence, one would be incoherent. But we may still wonder: why should we conceive of IBE *that* way? Certainly, when thinking about IBE, one *broad* way to model it as a rule is to cast it within a probabilistic (Bayesian) framework. This can be contested (cf. my 2004). But even if it is granted, there are still ways to show that van Fraassen's argument against the probabilistic casting of IBE is problematic. Two points are relevant here. The first, made by Jonathan Kvanvig (1994: 338) is that the way IBE is described by van Fraassen is unfair. Peter is portrayed as having an *inconsistent* attitude towards the very rule he is supposed to abide by, namely, IBE. For Peter ignores this rule *before* he learns the evidence *e* (he uses, instead standard Bayesian techniques), but he employs this rule *after* he has learned *e*. As Kvanvig says: 'what must be shown is that a consistent application of an IBE strategy is subject to Dutch Book difficulties' (1994: 338). But this has *not* been shown. If Peter followed IBE consistently, he would also have to change his prob(H) after having learned that H is the best explanation of the evidence *e*. If he did so, he wouldn't be subject to a Dutch Book strategy any more. Nor is it required for Peter to announce his belief-change strategy in advance. The second point, made by Gilbert Harman (1999: 110−11), is that if one were to render IBE within a Bayesian framework, then the explanatory considerations in favour of a hypothesis that is said to be the best explanation of the evidence should be reflected in its *prior* probability. If so, it is no longer the case that a Dutch Book strategy can be followed against an advocate of IBE.

Peter could follow a probabilistic version of IBE consistently if *either* of the following happened: a) the fact that a certain hypothesis H was the

152 / **Putting a Bridle on Irrationality**

best explanation of the evidence *e* was reflected in how Peter assigned *prior* probabilities to the competing hypotheses; b) the fact that a certain hypothesis *H* was the best explanation of the evidence *e* made Peter re-adjust both his prob(H/e) *and* his prob(H) too.[22] These are genuine options, if one wanted to render IBE within a probabilistic framework. Following IBE is not *ipso facto* incoherent.

Yet, there is another *broad* way to think of IBE as a rule of ampliative, and hence, defeasible reasoning, which is disconnected from any attempt to cast IBE within a probabilistic framework. It involves thinking of defeasible reasoning in general within a framework that connects justification with the absence of defeaters (see Pollock 1986). I have tried to analyse this approach in my (2002). Here is a summary statement of the view.

Unlike deductive methods, ampliative methods (IBE in particular) are *defeasible*: further information can remove the *warrant* for holding the output of the method. Following Pollock, we can call 'prima facie' or 'defeasible' any type of reason which is not conclusive in the sense that it is not deductively linked with the output it is a reason for. Given that ampliative reasoning is defeasible, we can say that such reasoning provides *prima facie warrant* for an output (belief). What Pollock has rightly stressed is that to call a warrant prima facie is not to degrade it, qua warrant or reason. Rather, it is to stress that a) it can be defeated by further reasons (or information); and b) its strength, qua reason, is a function of the presence or absence of 'defeaters'. 'Defeaters' are factors (generally, reasons or information) such that, when they are taken into account, they can remove the prima facie warrant for an outcome (belief). The presence or absence of defeaters is directly linked to the degree in which one is warranted in holding a certain belief. To say that *S* is prima facie warranted to accept the outcome *Q* of an ampliative method is to say that although it is possible that there are defeaters of the outcome *Q*, such defeaters are not actual. In particular, it is to say that *S* has considered several possible defeaters of the reasons offered for this outcome *Q* and has shown that they are not

---

[22] In his excellent (1999), Douven suggests, in effect, a combination of (a) and (b). His point is that IBE can be defended as a non-Bayesian rule of updating, against van Fraassen's argument, provided this happens: the conditional probability of a hypothesis given some piece of evidence is fixed in advance in such a way that it corresponds to the rule used to update degrees of belief when the evidence rolls in. Specifically, if a hypothesis that is taken to be the best explanation of the evidence gets a bonus increase of its posterior probability, then this bonus *should already have been reflected* in the calculation of the probability of the hypothesis conditional on that evidence.

present. If this is done, we can say that there are no *specific* doubts about the outcome of the method and that belief in this outcome is prima facie warranted. When it comes to IBE, the issue is to specify the possible defeaters (for example, competing explanatory hypotheses) of a conclusion licensed by an inference to the best explanation, and to examine whether they obtain in the particular case. We can conclude then that van Fraassen has not offered us compelling reasons to think that IBE is incoherent. IBE, as a rule, is *not* outside the bounds of reason.

But van Fraassen's agenda is broader. He aims to show that rule-following is not *required* for rationality. Suppose this is granted. Suppose, that is, that following IBE is not rationally compelling: one may refrain from accepting a hypothesis on the basis that it is the best explanation of the evidence. What follows from this? Not a lot. If IBE does not fall outside the bounds of reason, then one can reasonably follow it. What more can a defender of IBE ask for? The friends of IBE would be in great difficulty had it been shown that IBE was incoherent or irrational. But, as noted above, this has not been shown. It is noteworthy that Pollock too takes justification to be a matter of 'epistemic permissibility' (1986: 124) and subject to 'epistemic norms', where these 'norms describ[e] when it is epistemically permissible to hold various beliefs' (1986: 124−5). Epistemic norms tell a subject what she is *permitted* to believe under certain circumstances. In this setting, Inference to the Best Explanation could function as such a norm: it describes the circumstances under which it is rationally permissible to hold a belief based on explanatory considerations.

In any case, evaluativism is consistent with rule-following. One rather subtle issue that crops up especially in van Fraassen's critique of naturalized epistemology concerns the *justification* of ampliative rules. Van Fraassen seems to express the view that attempts to vindicate ampliative rules will presuppose these ampliative rules and hence that they can neither support nor undermine these rules (see especially 1995b: 78−81). In other words, it seems that van Fraassen points to the fact that the justification of ampliative rules would be *circular*. In my (1999: 81−90), I argued that some circular defences of basic inferential principles are epistemically significant. In particular, I argued that a) there is a difference between premise-circularity and rule-circularity (a premise-circular argument uses its conclusion as one of its premises; a rule-circular argument employs the *rule* it vindicates); b) rule-circularity is not vicious; and c) the circularity involved in the defence of basic rules of inference is rule-circularity. Though these points had already been made with regard

154 / **Putting a Bridle on Irrationality**

to basic deductive and inductive rules, I showed how a rule-circular defence could be offered on behalf of IBE. I am not going to repeat this defence here (but see also Foley 1993: 76−7). What I want to note is that this defence is fully compatible with evaluativism. Put in a nutshell, the (threefold) point is this. One: evaluativism makes plain that any attempt to justify a rule (ultimately by a rule-circular argument) will be an attempt for rules *we* value and will depend on rules we value (our basic inferential rules). Two: one of the things we value in our rules is their *reliability*, while another thing is that they are our *own* rules. Three: there is still space for comparative judgements. We can still assess a method in terms of its reliability to produce some results and we can still say that some methods are better than others in achieving a certain goal.

Here again, I am not saying that van Fraassen would necessarily deny all this. Yet, it is important to stress that evaluativism and rule-following mix well. In fact, casting the issue within evaluativism shows how and why the reliance of IBE on explanatory *virtues* is not damaging to the rationality of IBE. It's not just that it is permitted to rely on these virtues. It is also that these virtues are valued by us and can be linked with the truth-conduciveness of IBE.[23]

I think van Fraassen is very right if he means to stress that rational judgements (and what we call the scientific method, in general) are not *algorithmic*. He is also very right if he means to stress that there are no rules that leave nothing to one's choice. This, in fact, is the central message of evaluativism. But, if anything, it is a chimera to look for such rules. Duhem again brought this point home back in 1906.

6. The Bounds of Reason

It's now time to focus our attention on van Fraassen's (2000: 277) *central dictum*:

what is rational is precisely what is rationally permitted. Thus we are rational in believing something exactly when we are not rationally compelled to believe the opposite.

The concept of rational compulsion does enter van Fraassen's account, if only negatively: there must be some things that we are rationally compelled to believe or disbelieve. Otherwise, the idea of rationality as permission to believe would be vacuous. A lot, then, depends on how rational compulsion

---

[23] For more on this see my (1999: 171−6) and McMullin (1996).

is meant—what its scope is. Here is how van Fraassen (ibid.) elaborates on his position:

This [the central dictum] implies, tautologically, that *nothing more* than staying within the bounds of reason is needed for this status of rationality—not good reasons, not a rationale, not support of any special sort, not a pedigree of inductive reasoning or confirmation, nothing is needed above and beyond coherence. Thus any truly coherent position is rational.

Coherence is both deductive and probabilistic (synchronic); but nothing more than that. So what needs to be looked at is this: is the bound on reason captured by considerations of coherence only?

This is a huge issue. Is coherence (in both senses) necessary for rationality? Perhaps, it can be instituted as a *normative* constraint. Even this is debatable. The concept of rationality should be applicable to real agents. But it's hard to demand that real agents be coherent. To put it differently, if coherence is necessary for rationality, then real agents are irrational! Harman has persuasively argued that the deductive consistency of a set of beliefs is not necessary for rationality (1999: 18—20). Briefly put, his point is that demanding consistency among one's beliefs is an *unrealistic* constraint. Consistency is not, always, easy to establish. And one is not irrational if one cannot do what *cannot* be done. Of course, discovering an inconsistency in one's web of belief calls for its removal. But it is not always clear which belief should be removed. There may be reasons supporting each of the mutually inconsistent beliefs. It's not clear, to say the least, that one is irrational if one fails to remove the inconsistency, or if, having more urgent matters to attend to, one decides to live with it, at least for the time being. Things are equally bad when it comes to demanding (synchronic) probabilistic coherence as a necessary condition for rationality. Strict coherence requires that the subject has a degree of belief 1 in all and only noncontingent truths. Otherwise, the subject is open to a Dutch Book. But why, as Plantinga (1993) asks, should this be a requirement for rationality? It is perfectly rational that a subject has less than full degrees of belief in noncontingent propositions. And it is perfectly rational for a subject who knows that her degrees of belief are incoherent to avoid betting with a logically omniscient bookie (that is, with a bookie who assigns probability 1 to all and only noncontingent truths).[24]

[24] For a sustained critique of the view that rationality requires coherence, see Foley (1993: 155—84).

156 / **Putting a Bridle on Irrationality**

Perhaps, van Fraassen takes rationality to be a concept that applies to an *ideal* agent. And it is hardly an option to say that an ideal agent need not be coherent. Let me grant that coherence (in both senses) is necessary for rationality. Is it sufficient? I will argue that it is not. But let me first make clear how I perceive the dialectic of the situation. Van Fraassen claims that an agent is rational if she is deductively and probabilistically coherent. It then is enough to argue against this view that there is at least one clear-cut case in which an agent is deductively and probabilistically coherent and yet she is irrational. Calling a case 'clear cut' might be a rhetorical device. I don't deny this. I think we cannot leave our intuitions behind when we think about rationality. I will appeal to the reader's intuitions, knowing that they might not be enough to clinch the issue. In a sense, my point will be that mere coherence makes it too easy to be rational. This might be taken to be an advantage of van Fraassen's account. But it has an odd consequence: it goes against the deep-seated intuition that rationality has to do with what an agent does to make sure that her beliefs make contact with the world. From the agent's own perspective, this attempt to make contact with the world is connected with how the agent takes account of the evidence there is for her beliefs.

Here is my argument. Take someone who believes that the earth is flat. Indeed, there is a Flat Earth Society, with a page on the Internet, a theory why the earth is flat (actually claiming that it is shaped in the form of a pentagon) and a number of answers to frequently asked questions, explaining how things appear otherwise. The belief-corpus of the flat earthers is (or can be with enough ingenuity) deductively consistent and probabilistically coherent. Yet, this belief-corpus is irrational. Hence, there must be more to rationality than coherence.

This argument presupposes that the flat earthers are irrational. This may be contested. Here then is an argument to show why they are irrational: it is irrational to disregard evidence which is relevant to the truth or falsity of one's beliefs. The flat earthers disregard this evidence. Ergo, they are irrational. I take it as obvious that there is a lot of evidence showing that the earth is round and very little (if any) evidence showing that it is flat. The sufficiency of coherence allows the advocates of flat earth to disregard this evidence. In fact, they can find refuge in their being coherent, thereby legitimizing why they disregard relevant evidence. But this cannot be right. So there must be more to rationality than coherence.

Two objections may crop up at this point. The *first* is this. Suppose someone says that prob(H/e)=prob(H), where H is the hypothesis that the earth is flat and e is some potentially undermining evidence. The objection then is that the advocates of flat earth do not disregard the evidence. They just deem it irrelevant to the truth of H: they claim that e has nothing to say about H. My reply to this objection is that it concedes more than it intends to. The objection intends to show that flat earthers do not disregard relevant evidence. But in order to do so, it concedes that the evidence should *not* be disregarded! This is already a substantive principle of rationality that goes beyond coherence. It's irrelevant, at this stage, how this principle of rationality might be implemented. One might say that one has taken all evidence into account, or one might say that some of it is irrelevant to one's beliefs. Both attitudes require reliance on the more substantive principle of rationality.

Here is the *second* objection. One might argue as follows. This so-called substantive principle of rationality is really *empty*. For unless it is specified what counts as evidence for a belief, when the evidence is relevant and when not and so on, there is no content in the dictum: an agent shouldn't disregard the evidence for her beliefs. Interestingly enough, the present objection can find some solace in van Fraassen's own work. In his (1985), van Fraassen talks extensively about the 'judgements of evidence' and argues that these judgements presuppose 'criteria of relevance and judgements of comparison [which] are not written in the evidence' (1985: 278; cf. also 2002: 87). Moreover, he takes these criteria and standards to be context-dependent (1985: 279). It might then be open to flat earthers to argue that they simply have different criteria of relevance and different standards of comparison. My own reply to this objection proceeds in two steps. The first step is to grant that the criteria of relevance are not written in the evidence and can be context-dependent. Actually, I think van Fraassen is right in criticizing this 'revelation model of the evidence' (cf. 1985: 250; 1980: 168). Surely, the evidence does not (always) speak with the voice of an angel (cf. 1980: 169). The second step is this: it does not follow from the above that it is always rational for an agent to disregard some evidence either by simply denying that this is evidence or by appealing to her own criteria of relevance or to the context. In fact, most people (even the flat earthers) do not. Some pieces of evidence constitute empirical truisms that cannot be coherently denied. The creationist won't deny that he sees fossils. And the flat earther won't deny that he sees photographs in which the earth appears to be round. Some relevance relations are fixed for all sides of the

158 / **Putting a Bridle on Irrationality**

debate. The creationist won't deny that fossils are relevant to his creationist stories. And the flat earther won't deny that the photographs of the earth are relevant to his flat earth story. The onus is on them to show why they are entitled to accept some things as evidence or as criteria of relevance but not others. In fact, something stronger can be asserted. They cannot appeal to different criteria of relevance in order to jettison some piece of evidence as irrelevant. For in order to talk about their criteria of relevant evidence being different from their opponents', they need to appeal to some *evidence*. If *this* evidence is relative to some criteria of relevance, then they are faced with the well-known dilemma: either they will beg the question or they will be involved in infinite regress.[25]

The claim that an agent shouldn't disregard the evidence for her beliefs is not an empty dictum. A rational agent should regard *all* evidence that bears on a certain belief (or hypothesis) judiciously, try to take it into account in coming to adopt a belief (or a hypothesis) and then form her judgement in its light.[26] This principle (let's call it *the principle of evidential support*) goes far beyond the demand of coherence. It is a substantive principle of rationality. In fact, it is necessary for rationality on pain of not allowing one's beliefs to make any contact with the world. Certainly, if the evidence is not enough to clinch an issue, agnosticism might be the right attitude. But it is hardly the case that agnosticism is an option with respect to a host of issues.

It is not a reply to the argument above that the evidence can and does have a bearing on how beliefs *change*. This should of course be granted. But this is another substantive principle of rationality that goes *beyond* coherence. Someone who takes coherence to be sufficient for rationality needs an extra principle in order to make evidence *count* in belief revision.

Here is another worry about van Fraassen's voluntarist conception of rationality. Take a proposition *p* such that it is rationally permitted for an

---

[25] Don't scientists disregard some evidence some time? Of course they do. But there is a relevant difference here. This is that scientists disregard some negative evidence (say Kuhnian anomalies) for a good *reason*—viz., that the theory that faces the recalcitrant evidence has had many predictive and explanatory successes, which warrant the belief that there may be a successful incorporation of the recalcitrant phenomena. This might never happen. But that's exactly the point. Even if it may be reasonable to disregard some evidence for some time, it is not reasonable to disregard some evidence forever. The weight of negative evidence helps (among other things) to sway the advocates of one theory to another.

[26] It might not be irrational for an agent to disregard the evidence if some important non-epistemic goals are at stake. But this is hardly an option if the goal is epistemic.

agent *S* to believe that *p* because *S* is not rationally compelled to believe the opposite (that is not-*p*). If no substantive principles are involved in determining what an agent is rationally *compelled* to (dis)believe, and if, in particular, belief in any non-contradictory proposition is rationally permitted, then *S* will also be rationally permitted to believe that not-*p*, since *S* will *not* be rationally compelled to believe that *p*. Van Fraassen's conception of rationality will allow that both belief in *p and* belief in not-*p* be rationally permitted. This will be the case for most empirical propositions, unless belief in some of them and/or disbelief in others is rationally compelling. Now, what I have just noted is not necessarily a problem. Actually, it might be taken to be one of the prime attractions of van Fraassen's views that it makes all this possible. However, this situation has two interesting consequences, which support the view that there is more to rationality than coherence.

The first is that a voluntarist agent (let's call her Mary) can decide to live with contradictory beliefs. Since belief in either *p* or not-*p* is rationally permitted, Mary can display belief in *p* in some contexts (maybe on Mondays, Wednesdays, and Fridays) and belief in not-*p* in some other contexts (maybe on Tuesdays, Thursdays, and Saturdays). She should, of course, refrain from displaying both beliefs at one time simultaneously. But with enough care, she can do this. (On Sundays, Mary is agnostic.) There is something wrong with Mary's belief system. But it's not clear that van Fraassen's conception of rationality can show what it is. If coherence is sufficient for rationality, then Mary can divide her belief systems into two sub-systems, one of which includes belief in *p* while the other includes belief in not-*p*. Given the obvious ad hocness of this move, there must be more to rationality than coherence.

The second consequence is this. If both belief in *p* and belief in not-*p* are rationally permitted (since neither belief in *p* nor belief in not-*p* is rationally compelling), Mary can use each belief to undermine the other. Since it is rationally permitted for her to believe in *p*, she can use this as a reason to disbelieve not-*p*. But since it is also rationally permitted for her to believe in not-*p*, she can use this as a reason to disbelieve *p*. So she has a reason not to believe in *p* and a reason not to believe in not-*p*. She will then end up with no belief on the matter. If we now suppose that a situation like this may occur for any contingent proposition *p*, Mary may end up with no beliefs at all. Worse, she may end up refraining from believing in anything. In fact, this supposition is not implausible within van Fraassen's framework of rationality. The *central dictum* is too thin. Unless there are substantive principles which make disbelief

in some propositions rationally compelling, belief in them will be permitted. But so will be belief in their opposites. Mary will end up in a Buridan's ass kind of situation. She will remain undecided as to what she is to believe.

Perhaps, this is where voluntarism has a bite. Mary will just *decide* what to believe by a kind of leap of faith. She might well say that belief in, say, not-*p* is not a 'live option' for her (cf. van Fraassen 2002: 99). But it will be odd, to say the least, if, while she made this leap, she *knew* (as she does) that she could have been equally rational if she had made a leap of faith in the opposite direction. 'I believe that *p* but I could equally well have believed that not-*p*' may not be incoherent, but it seems to give rise to a deep intellectual tension. Then, an issue with van Fraassen's voluntarism is how we (should) choose between rationally permitted beliefs. In his (2002: 97) he says (agreeing with Pascal) that 'there are still distinctions to be drawn among beliefs when neither they nor their contraries are compelled by reason and evidence.' But drawing these distinctions requires going beyond coherence (both deductive and probabilistic).

It might be thought that talking of standards of rational belief (or judgement) implies the existence of some topic-neutral criteria. This need not be so. Consider the case of clinical trials. There is a very detailed methodology that has to be followed if the outcome of a clinical trial is to be reliable. Judgement based on the results of clinical trials (for example, that a certain drug is to be administered to those who suffer from depression) is made *rational* by the fact that a certain method has been followed that has met certain domain-specific standards. These are objective judgements. Their results can be independently checked and the experiment can be repeated. The adequacy of the method can be justified. The method can be compared with others and be defended as a better one. But this method determines what is rational to believe and what is not. Suppose that a drug has been shown to be effective against depression by a well-conducted clinical trial. There is still space for disagreement, of course. One may doubt the design of the trial. Another may claim that some confounding variables were not controlled for and so on. But if it is made sure that none of these objections hold water, it is no longer rational to doubt the result. Maybe it is *still* rational for a philosopher; but not for the working scientists. Here van Fraassen's dictum that 'we are rational in believing something exactly when we are not rationally compelled to believe the opposite' (2000: 277) might be found wanting. For if scientific methodology (for example, of the form of clinical trials) is taken into account

when considering what we are compelled to disbelieve, then his dictum would betray the *thin* notion of rationality that he wants to defend. If, on the other hand, ordinary scientific methodology is not taken into account when considering what we are compelled to disbelieve, then this would follow: since considerations of coherence alone leave it open to believe in an outcome of a clinical trial or in its negation, both beliefs would be equally rational. But they need not be.

A reply that van Fraassen would offer might be this. In his (2000: 274–5) he takes the view that what has been called the scientific method is (almost) nothing more than logic and mathematics. He focuses on what he calls 'techniques' which 'are not bound to any special built-in assumptions or historical constraints' (2000: 275). He might then reply to the point above that what I called substantive scientific methodology is nothing more than highly sophisticated statistical methods. Hence, he could say, they belong to the realm of mathematics and, as such, they are 'empty' of content. This reply would be wrong, however. For though the methodology of clinical trials *does* contain a set of sophisticated statistical techniques, it does not contain *just* them. It rests on certain substantive principles (randomization is one of them). More to the point, it is well-known that the method of double-masked experiments was developed after certain substantive discoveries were made about the spontaneous recovery of subjects that knew they were, say, administered a drug. The development of these 'techniques' was the joint product of fine developments in statistics and substantive discoveries about the world (largely based on ampliative reasoning—and on IBE in particular, I should add). Their legitimacy rests not only on the correctness of the statistical methods but also on substantive (and mostly ampliative) beliefs about the world.

7. Concluding Thoughts

Van Fraassen is right that 'irrationality' is, so to speak, the primary notion. But I disagree that attributions of irrationality pertain only to the structure of a belief-corpus and (possibly) to belief-change. I side with Nozick's (1993: 87) view that 'it is clear that many things are irrational to believe.' Irrationality pertains to the *content* of a belief too—perhaps primarily so. And at least *some* ascriptions of irrationality to belief-contents seem both compelling and inescapable. Yet, I also agree with Nozick's view that 'it is less clear that some beliefs are so credible

162 / **Putting a Bridle on Irrationality**

that they are mandated by rationality, so that it is irrational not to hold them when you hold no belief about the matter at all' (ibid.) Van Fraassen should be credited for strengthening this point. He is right in saying that rationality involves *permission*. Indeed, it is rationally permitted to believe in many things: in doing so we don't flout any criteria of rationality. Yet, rationality also involves *obligation*. Some criteria of rationality (mostly substantive: beliefs should be based on evidence or beliefs should be formed by reliable means or methods) set down obligations: that it is not rationally permitted to believe (at least any more) in many things, because so believing would flout some criteria of rationality.

What, I think, is right with van Fraassen's new epistemology is its attempt to deflate the concept of rationality. But I also think it's wrong in its attempt to do away with all substantive criteria of rationality. Purely formal criteria (deductive and probabilistic coherence) are not sufficient for rationality. (And, arguably, they are not necessary either for a non-idealized conception of rationality.) Purely formal criteria of belief-revision (such as Bayesian conditionalization) are very limited—they are not applicable across the board and, as van Fraassen says, they are not mandatory even where they apply. Substantive rules (such as inference to the best explanation) are not inconsistent. They are not, to be sure, rules suitable for a Carnapian robot, but this is as it should be. The attempt to offer an algorithmic conception of substantive rules (and of scientific method, in general) has been a great failure. And it is an error to try to stay on this course. The algorithmic conception of rationality is nothing more than a phantom. Whatever else it does, the scientific method relies essentially on background knowledge about what the world is like and requires the exercise of judgement. Consequently, the scientific method is not *just* a set of 'empty' logico-mathematical techniques.

Recall that van Fraassen says that rationality is but bridled irrationality. This seems correct. But I think the real issue between van Fraassen's conception of rationality and opposing views concerns the 'bridle': How substantive is it? Does it allow for rule-following? And so on. The bridle won't leave *nothing* to our choice. It won't exclude values. But it constrains choices in a way more substantive than van Fraassen seems to think.

Then, in the spirit of unimpeded and open-minded inquiry, we can have a rational deliberation over some central issues: Is the evidence enough to warrant belief in *x*? If it is less than enough, what else would be required? Could

some evidence *ever* be enough for a certain belief? How should we go about collecting more evidence? What methods shall we use? And in the same spirit, we can start pondering whether belief, disbelief, or agnosticism is the right attitude in the certain context.

Van Fraassen (2000: 279) summarizes his position thus:

We supply our own opinion, with nothing to ground it, and no method to give us any extra source of knowledge. Only the 'empty' techniques of logic and pure math are available either to refine and improve or expose the defects of this opinion. That is the human condition. But it is enough.

One may well wonder whether this is indeed the human condition.[27]

References

CHRISTENSEN, DAVID (1991), 'Clever Bookies and Coherent Beliefs', *The Philosophical Review* 100: 229−47.

DOUVEN, IGOR (1999), 'Inference to the Best Explanation Made Coherent', *Philosophy of Science (Proceedings)* 66: S424−S435.

DUHEM, P. (1906), *The Aim and Structure of Physical Theory* (Princeton: Princeton University Press).

EARMAN, JOHN (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (Cambridge MA: MIT Press).

FIELD, HARTRY (2001), *Truth and the Absence of Fact* (Oxford: Oxford University Press).

FOLEY, RICHARD (1993), *Working Without a Net* (New York: Oxford University Press).

HARMAN, GILBERT (1999), *Reasoning, Meaning, and Mind* (Oxford: Oxford University Press).

HOWSON, COLIN (2000), *Hume's Problem* (New York: Oxford University Press).

KVANVIG, JONATHAN (1994), 'A Critique of van Fraassen's Voluntaristic Epistemology', *Synthese* 98: 325−48.

LANGE, MARC (1999), 'Calibration and the Epistemological Role of Bayesian Conditionalisation', *Journal of Philosophy* 96: 294−324.

164 / **Putting a Bridle on Irrationality**

McMullin, Ernan (1996), 'Epistemic Virtue and Theory Appraisal', in I. Douven and L. Horsten (eds.), *Realism in the Sciences* (Leuven: Leuven University Press), 13−34.

Nozick, Robert (1993), *The Nature of Rationality* (Princeton: Princeton University Press).

Plantinga, Alvin (1993), *Warrant: The Current Debate* (Oxford: Oxford University Press).

Pollock, John (1986), *Contemporary Theories of Knowledge* (Savage MD: Rowan & Littlefield).

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth* (London: Routledge).

——— (2002), 'Simply the Best: A Case for Abduction', in F. Sadri and A. Kakas (eds.), *Computational Logic: From Logic Programming into the Future* (Berlin: Springer-Verlag), 605−25.

——— (2004), 'Inference to the Best Explanation and Bayesianism' in F. Stadler (ed.), *Induction and Deduction in the Sciences* (Dordrecht: Kluwer), 83−91.

Skyrms, Brian (1984), *Pragmatics and Empiricism* (New Haven: Yale University Press).

Teller, Paul (1973), 'Conditionalisation and Observation', *Synthese* 26: 218−58.

van Fraassen, Bas (1980), 'Rational Belief and Probability Kinematics', *Philosophy of Science* 47: 165−87.

——— (1983), 'Calibration: A Frequency Justification for Personal Probability', in R. S. Cohen and L. Laudan (eds.), *Physics, Philosophy and Psychoanalysis* (Dordrecht: D. Reidel Publishing Company), 295−319.

——— (1984), 'Belief and the Will', *Journal of Philosophy* 81: 235−56.

——— (1985), 'Empiricism in the Philosophy of Science', in P. M. Churchland and C. A. Hooker (eds.), *Images of Science* (Chicago: The University of Chicago Press), 245−308.

——— (1989), *Laws and Symmetry* (Oxford: Oxford University Press).

——— (1993), 'From Vicious Circle to Infinite Regress, and Back Again', in D. Hull, M. Forbes, and K. Okruhlik (eds.) *PSA 1992*, Vol. 2 (Evanston: Northwestern University Press), 6−29.

——— (1995a), 'Belief and the Problem of Ulysses and the Sirens', *Philosophical Studies* 77: 7−37.

——— (1995b), 'Against Naturalized Epistemology', in P. Leonardi and M. Santambrogio (eds.), *On Quine* (Cambridge: Cambridge University Press), 68−88.

——— (2000), 'The False Hopes of Traditional Epistemology', *Philosophy and Phenomenological Research* 60: 253−80.

——— (2001), 'Constructive Empiricism Now', *Philosophical Studies* 106: 151−70.

——— (2002), *The Empirical Stance* (New Haven: Yale University Press).

Williams, Bernard (1973), *Problems of the Self* (Cambridge: Cambridge University Press).

——— (2002), *Truth and Truthfulness* (Princeton: Princeton University Press).

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

Worrall, John (1993), 'Falsification, Rationality and the Duhem Problem', in J. Earman, A. Janis, G. Massey, and N. Rescher (eds.), *Philosophical Problems of the Internal and the External Worlds* (Pittsburgh: Pittsburgh University Press), 329−70.