Körner, Stephan (1970), *Categorial Frameworks* (Oxford: Blackwell).

Leeds, Stephen (1978), 'Theories of reference and truth', *Erkenntnis* 13: 111–29.

Locke, John (1690), *An Essay Concerning Human Understanding* (New York: Dover, 1959).

Maddy, Penelope (1997), *Naturalism in Mathematics* (Oxford: Oxford University Press).

Matthews, H. E. (1969), 'Strawson on transcendental idealism', *Philosophical Quarterly* 19: 204–20.

Pitcher, George (1977), *Berkeley* (London: Routledge & Kegan Paul).

Prichard, H. A. (1909), *Kant's Theory of Knowledge* (Oxford: Oxford University Press).

Putnam, Hilary (1971), 'Philosophy of Logic', repr. in his *Philosophical Papers*, ii, 2nd edn. (Cambridge: Cambridge University Press, 1979): 323–57.

—— (1981), *Reason, Truth and History* (Cambridge: Cambridge University Press).

Quine, W. V. (1948), 'On what there is', repr. in his (1980): 1–19.

—— (1951*a*), 'Two dogmas of empiricism', repr. in his (1980): 20–46.

—— (1951*b*), 'Carnap's view on ontology', repr. in his (1976): 203–11.

—— (1955), 'Posits and reality', repr. in his (1976): 246–57.

—— (1960), *Word and Object* (Cambridge, Mass.: MIT Press).

—— (1970), *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall).

—— (1975), 'Five milestones of empiricism', repr. in *Theories and Things* (Cambridge, Mass.: Harvard University Press): 67–72.

—— (1976), *The Ways of Paradox*, revised edn. (Cambridge, Mass.: Harvard University Press).

—— (1980), *From a Logical Point of View*, 2nd edn. (Cambridge, Mass: Harvard University Press).

—— (1981*a*), 'Things and their place in theories', in (1981b) 1–23.

—— (1981*b*), *Theories and Things* (Cambridge, Mass: Harvard University Press).

Reichenbach, Hans (1920), *The Theory of Relativity and A Priori Knowledge*, trans. M. Reichenbach (Berkeley and Los Angeles: University of California Press, 1965).

—— (1949), 'The Philosophical significance of the theory of relativity', in P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist* (La Salle, Ill.: Open Court): 287–311.

Reichenbach, Maria (1965), 'Introduction to Reichenbach' (1920): pp. xi–xliv.

Strawson, P. F. (1966), *The Bounds of Sense* (London: Methuen).

Van Fraassen, Bas (1980), *The Scientific Image* (Oxford: Oxford University Press).

6

Apriority as an Evaluative Notion

Hartry Field

A priori justification is often thought mysterious or out of keeping with a naturalistic view of the world; strong forms of a priori justification that involve empirical indefeasibility are often thought especially mysterious. While this is no doubt correct for *excessive* claims of a priority—for instance, claims to a priori access to features of the physical world—I will argue that it is incorrect if intended as a claim about the existence of any apriority at all.[1] What is mysterious in most forms of (non-excessive) apriorism isn't the apriorism itself but the background assumptions about epistemology. But in questioning these background assumptions, I will be producing an account of apriority that few apriorists will like.

1. THE CONCEPT OF APRIORITY

Let's define a *weakly a priori* proposition as one that can be reasonably believed without empirical evidence;[2] an *empirically indefeasible* proposition as one that admits no empirical evidence against it;[3] and an *a priori* proposition as one that is both weakly a priori and empirically indefeasible. Some writers use 'a priori' in a way that imposes no empirical indefeasibility requirement, but it seems to me that that removes the main philosophical interest of apriority: traditional debates about the apriority of logic and Euclidean geometry have largely concerned the issue of whether any empirical evidence could count against them. Another reason for keeping an indefeasibility requirement will be given later in this section.

Thanks to Ned Block, Paul Boghossian, Stephen Schiffer, and Nick Zangwill for extremely helpful discussions.

[1] The mystery that excessive claims to apriority would create is briefly discussed in n. 19. I believe that there is no analogous mystery for apriorism about logic, or for apriorism about the basic features of scientific method.

[2] Here and throughout, 'reasonable' will mean 'epistemically reasonable': crassly pragmatic motivations for and against believing are not to be taken into account.

[3] 'Empirically indefeasible' may be too weak a term for what I've just defined: the term suggests only that there can never be *sufficient* empirical evidence against it to outweigh any initial plausibility it might have. But it isn't easy to find examples that meet the weaker condition but not the stronger, and I will continue to use the term in the stronger sense.

The empirical indefeasibility requirement does need to be either restricted or interpreted delicately if it is not to immediately rule out a priori knowledge. As a first illustration (emphasized in Kitcher 1983), the credibility of any proposition could be diminished by evidence that well-regarded experts don't accept it. This first illustration doesn't seem to me very interesting: perhaps it shows that we must impose a slight restriction on empirical indefeasibility to allow for defeasibility by the opinions of others,[4] but surely it doesn't suggest that we should give up an indefeasibility requirement entirely (so that we could reasonably regard a proposition of logic as a priori while at the same time granting that experimental results in quantum mechanics could tell against it).

But there is a more interesting argument against an empirical indefeasibility requirement. The argument has two steps. First, empirical indefeasibility seems equivalent to *empirical unaugmentability*: the condition that there can be no empirical evidence *for* the proposition. Their equivalence follows from the hard-to-contest principle that an experience counts as evidence for a proposition only if some contrary experience would count as evidence against it and vice versa. But second, as has often been noted, complex and unobvious logical truths can admit empirical justification without diminishing their claims to a priori status. For instance, a proposition of form $((p \supset q) \supset p) \supset p$ is obviously entailed by p, so someone who didn't realize it was a logical truth might empirically justify it by empirically justifying p. So a proposition that should be an extremely plausible candidate for apriority seems empirically augmentable and therefore (given the equivalence) empirically defeasible.

The best way to deal with this argument, is to distinguish empirical justification and empirical evidence: evidence involves something like *ideal* justification, ideal in that limitations of computational capacity are ignored. The idea is that reflection on the logical facts reveals that the evidence for p doesn't raise the 'ideal credibility' of the logical truth $((p \supset q) \supset p) \supset p$: for ideally that would have been fully credible to begin with. If an observation doesn't raise the *ideal* credibility of the claim, it shouldn't count as evidence for it. Similarly, an observation must lower the *ideal* credibility of a claim to count as evidence against it. A nice thing about this resolution of the second argument against an empirical indefeasibility requirement is that it could be employed in the Kitcher example too: while the non-ideal credibility of, say, a complex logical truth can certainly be lowered by empirical evidence that well-respected logicians didn't accept it, ideal credibility can't be lowered in this way: for that reason, the evidence about the opinions of logicians really isn't *evidence* against the logical truth. Whether the Kitcher examples are to be handled this way or by a slight restriction on the empirical indefeasibility requirement is something I leave to the reader.

I want to say a bit more about my proposed definition of apriority, but first it would be well to generalize: it is important to consider not only the apriority of propositions, but the apriority of methodologies or rules for forming and revising beliefs. For the moment, we can take as examples of such methodologies (classical) *deductive inference*, (your favourite version of) *inductive inference*, and (your favourite rules of) *percentual belief formation*, i.e. of the formation of perceptual beliefs on the basis of experience. In analogy to the above, I will call such a methodology or rule *weakly a priori* iff it can be reasonably employed without empirical evidence; *empirically indefeasible* if no empirical evidence could undermine the reasonableness of its employment; and *a priori* if it meets both conditions. Again, I think the most interesting component of apriority is empirical indefeasibility.

Note that I have not required that an a priori proposition can only be reasonably believed by someone who has a non-empirical justification for it: not only would that conflict with the examples above of a priori propositions reasonably believed entirely because of empirical justifications, it would also conflict with the possibility of a priori propositions reasonably believed without any justification at all. ('Default reasonable' propositions.) Similarly in the case of rules. I think that we ought to allow for the possibility of default reasonable propositions and rules;[5] more on this shortly. My definition classifies default reasonable propositions and rules as, trivially, *weakly* a priori; so that they are a priori if and only if they are empirically indefeasible. If one were to hold that a priori justification is required for reasonable belief in an a priori proposition and for reasonable employment of an a priori rule, then default reasonable propositions and rules could never count as a priori. That would be most undesirable: surely among the most plausible examples of default reasonable propositions and rules are simple logical truths like 'If snow is white then snow is white' and basic deductive rules like modus ponens and 'and'-elimination. It would be odd to exclude these from the ranks of the a priori merely because of their being default reasonable.[6]

[4] One restriction that would have this effect was suggested in Field (1996).

[5] One must be careful not to be led into ruling them out by pun on the word 'justified'. In one sense, a justified belief is simply a reasonable belief; in another, it is a belief that has a justification. If it is assumed that these senses are equivalent, the exclusion of default reasonableness is automatic; but in fact their equivalence needs argument. (Note that if their equivalence is not assumed, there is no reason not to suppose that 'unjustified' beliefs in the second sense can be essential ingredients in the justification of other beliefs.)

[6] The problem of default reasonable propositions and rules is curiously overlooked in discussions of how the notion of a priori proposition and/or a priori justification is to be defined. For instance, the discussion of a priori justification in Bonjour 1998 assumes throughout that for a belief to be reasonable it must have some justification or other, if not empirical then a priori. Presumably Bonjour thinks that there are no default reasonable propositions. However, the obvious way to retain that position is to allow for circularity in the justificatory process (see the next section), and Bonjour makes a point of disallowing such circularity: that is the basis on which he argues that for induction to be reasonable it must be possible to give a justification of it that doesn't use induction. He thinks that such a non-circular justification of induction is possible; I will not discuss this, but unless he also thinks that a non-circular justification of deduction is possible, then the exclusion of circularity would seem to make the recognition of default-reasonable rules of deduction mandatory if deductive scepticism is to be avoided. (I suspect that

If our concept of apriority were simply weak apriority we would have the opposite problem: default reasonable propositions would automatically count as a priori. But there is no obvious reason why propositions such as 'People usually tell the truth' shouldn't count as default reasonable, and it would be odd to count such propositions a priori. Empirical indefeasibility seems the obvious way to distinguish those default reasonable propositions that are a priori and those that aren't.

There is another possibility worth considering: I have argued against saying that a priori propositions and rules are those that *require* non-empirical justification to be reasonably believed, but why not say that they are those that *admit* non-empirical justification? The answer is that this too might exclude simple logical truths, or rules like modus ponens and 'and'-elimination. For the only obvious way to try to give 'a priori justifications' for them is by appeal to the truth-table for 'and'. But as has often been pointed out, 'justification' of 'and'-elimination by the truth-table for 'and' requires the use of 'and'-elimination (or some equivalent principle) at the meta-level: one must pass from '"A" and "B" are both true' to '"A" is true'. If this counts as a justification it is a circular one,[7] and it is not obvious that 'circular justification' makes sense. I'll soon discuss that issue more fully, but at least we can say that the alternative approach to defining a priority contemplated at the start of this paragraph requires the acceptance of circular justification.[8]

I close this section by noting that it is not built into the definitions that an a priori proposition be true or an a priori methodology reliable; much less, that its truth or reliability is somehow *guaranteed* by some non-empirical justification of it. We do have strong *reason to think* that a prior propositions are true and a priori methodologies reliable: if we didn't have reasons to think these things, it wouldn't be reasonable to believe the propositions or employ the methodologies, so they couldn't be a priori.[9]

2. *DEFAULT* REASONABLENESS

There is a familiar argument for the default reasonableness of certain methodologies, including deductive reasoning, inductive reasoning, forming beliefs as a result of observation or testimony or memory-impression, and so forth. (Recall that if they are default reasonable then they are at least *weakly* a priori.) The argument is that no justification for anything could be given without using some of these forms of reasoning.[10] So if justifications are assumed to be non-circular, and if we exclude the totally sceptical possibility that no methodology for forming and revising beliefs is reasonable, then some methodologies must be reasonable without justification: they must be 'default reasonable'.[11]

Should we exclude all circular 'justifications' of methodological rules from being genuine *justifications*? A number of authors have argued against such an exclusion (Black 1958; Dummett 1978; Friedman 1979; van Cleve 1984), on what appear at first glance to be reasonable grounds. Indeed, at least part of what Dummett and Friedman say seems incontestable: a deductive justification of deduction does give us some kind of rational explanation of why we should value our deductive practice more highly than alternative deductive practices we consider defective. This is doubtless of importance—more on this later. But it is not obvious that its providing this sort of explanation of our valuing the practice means that it should count as a justification. To be sure, Dummett and Friedman grant that such circular explanations are not the sort of justifications that would persuade a committed proponent of alternative methods; but I take the issue of whether they count as justifications not to be that, but rather, whether they should add to the credibility we attach to the mode of reasoning in question. In my view, the explanations can have this justificatory value only if they aren't too easy to come by: only if there was a prima-facie risk of it being impossible to explain our valuing the method,[12] so that the actual providing of the explanation can justify the method by showing that the risk is not genuine. I think that in the case of deduction and induction and perception there is reason to doubt that there is a significant prima-facie risk, in which case it is hard to see why the circular 'justifications' should count as justifications at all. (More about this in the inductive and perceptual cases in Section 4.)

Even if we concede that such circular 'justifications' have justificatory value,

Bonjour would say that rather than being default reasonable, the basic rules of logic are justified by acts of a priori insight. But this seems like just an obscurantist redescription; in any case the only argument for it seems to rest on defining a priori justification in a way that ignores the possibility of default reasonableness.)

[7] 'Circular' here is taken to include 'rule-circular': the relevant sort of circularity is where we justify *the claim that a rule is truth-preserving* by use of that very rule.

[8] Relative to principles of justification which allow for circular justification, the alternative definition of apriority contemplated in this paragraph may be equivalent to the one I proposed.

[9] A further point worth mentioning: I do not assume that it is a failure of rationality to believe of a proposition that is not a priori that it is a priori, or to believe of one that is a priori that it is not.

[10] Admittedly, this might not be so if 'acts of a priori insight' are both possible and count as justifications; but let's agree to put obscurantism aside.

[11] This is compatible with 'externalist' views about reasonableness (as well as with 'internalist' views). The externalist holds that a necessary condition on the reasonable employment of inductive procedures or perceptual procedures is that those procedures in fact be 'reliable' or 'truth-conductive' or whatever (where the 'whatever' covers any other intuitively 'externalist' condition that might be imposed). This is compatible with certain procedures being default reasonable: it just implies (i) that what procedures are default reasonable depends on which ones satisfy the appropriate externalist conditions; and (ii) that *evidence in favour of* the satisfaction of those conditions isn't also necessary for the procedures to be reasonably employed. (I doubt that the contrast between internalist and externalist conditions is altogether clear, but I will not be making much of the contrast. In fact, I will eventually argue that even if that distinction is clear, the distinction between internalism and externalism rests on a false assumption.)

[12] Indeed, only if there was a prima-facie risk that in using our methods we will come to the conclusion that the methods do not have the properties we value.

there is a case for certain deductive, inductive, and perceptual rules being 'default reasonable'. Indeed, the case for default reasonableness is explicit in most of the works just cited: the authors argue that what makes the rule-circular justifications of certain rules count as justifications is that those rules already have a kind of initial credibility. (They think that use of initially credible rules to argue for the reliability or truth-preservingness of the rules adds to this initial credibility.) Their 'initial credibility' is my 'default reasonableness'.

It is, however, not out of the question to hold that without circular justifications there is no reasonableness at all. That is the view of a certain kind of coherence theorist. This coherence theorist holds that simple deductive, inductive and perceptual rules don't count as 'reasonable to employ' until the users of those procedures have argued (using some combination of deduction, induction, and perception, the combination varying from one case to the next) that those rules are reliable. But once the rules have been used to support themselves to a sufficient degree, the rules become reasonable to employ.

But I doubt that this way of avoiding the default-reasonableness of certain inferential rules has any substance. Presumably not any set of procedures that are self-supporting (i.e. which can be used in arguments for their own reliability) count as reasonable to employ: consider various sorts of counter-deductive and counter-inductive methods. What differentiates those which are reasonable (e.g. ours) from those which aren't? The natural answer is that our methods have a certain proto-reasonableness, independent of empirical evidence in their favour, that counter-deductive and counter-inductive methods lack. This proto-reasonableness might be due entirely or in part to factors like truth-preservingness or reliability; or it might be just due to the fact that we find these procedures natural to employ. Either way, once we use our method to reach the conclusion that that method is reliable, the proto-reasonableness is converted to full reasonableness; counter-deductive and counter-inductive methods don't have proto-reasonableness to begin with, so they don't become reasonable upon self-support. That, I submit, is the most straightforward way for a coherence theorist to differentiate the reasonable from the unreasonable self-supporting methods.

But then it is transparent that the view is basically just a notational variant of the view that there is default reasonableness; it just calls it proto-reasonableness. Of course, 'default reasonable' is supposed to imply 'reasonable', whereas 'proto-reasonable' is supposed not to imply it (and indeed, to imply that something else is needed before reasonableness is achieved); but my point is that the advocates of this view do ascribe a positive value to what they call proto-reasonableness, it's just that they adopt a higher threshold for the value that must be obtained to deserve to be called 'reasonable'.

There are two considerations that favour the lower (non-coherentist) threshold. First, if as I have suggested there is less than meets the eye to deductive justifications of deduction and inductive justifications of induction, the point of elevating the standard of reasonableness in the way the coherentist proposes is

obviously diminished. I'll say no more about this now. Second, I think that at least in the case of induction, it is impossible even using the rules in question to argue for crucial features of the reliability of the rules; this means that it is hard to motivate a higher (coherentist) threshold without motivating one so high that it is unattainable. Arguing this is a bit outside the main thrust of the paper, so I leave it to a footnote.[13]

Despite these considerations, the decision on 'threshold of reasonableness' is partly verbal, and this partly verbal decision affects the scope of weak apriority as I've defined it. Consider inductive and perceptual rules (good ones; this presumably includes the ones *we* use). On the lower (non-coherentist) threshold, such rules come out default reasonable and therefore weakly a priori. But on the higher (coherentist) threshold according to which good inductive and perceptual rules are merely proto-reasonable, then those rules don't count as weakly a priori unless they can be given non-empirical justifications; and this is most unlikely

[13] Here is the argument that empirical evidence for the reliability of relevant features of our inductive procedures is simply unavailable. Suppose we have developed a comprehensive and appealing physical theory T that the evidence at our disposal strongly supports. We can always invent bizarre alternatives that no one would take seriously, but which the available evidence equally accords with: for instance,

(T*) T holds at all times until the year 2000, at which time U holds

(where U is some detailed development of a totally discredited theory, say Aristotelian physics). The reason for saying that the available evidence accords with T* just as well as it accords with T is that the available evidence all concerns what happens at times before 2000, and T and T* agree completely about that. Despite this, we would of course all base predictions on T rather than on T*: it is part of our empirical methodology to do so, and surely doing so is reasonable. But it is hard to see that we have any evidence favouring this methodology over an alternative one which favours T* over T.

Someone might claim that we do have such evidence: the abundant evidence in our possession that the laws of physics haven't changed in the past. But this is a mistake: if the laws had changed in the past, that would be incompatible with both T and T*, so it wouldn't favour either over the other, and so evidence against it also doesn't favour one over the other. To make this clearer, let's look at two more theories besides T and T*:

V*: The current laws of physics are T; but the laws have changed every 100 years, and will continue to do so.

V: The current laws of physics are T; but the laws have changed every 100 years in the past. However, the laws won't change in 2000 or thereafter.

There is little doubt that if the laws had changed in the years 100, 200, ..., 1900, that would be pretty good inductive evidence that they would also change in 2000; and that the fact that they didn't change then is evidence that they won't change in 2000. The reason is that our methodology gives a strong a priori bias to V* over V and to T over T*. Evidence that the laws have changed in the past would rule out T and T* but leave V* and V as consistent with the evidence; given the a priori bias, V would be dismissed as highly implausible, leaving V*, which entails a change in the year 2000. Similarly, evidence that the laws haven't changed rules out V and V*, leaving T and T* as consistent with the evidence; but this time the a priori bias leads us to dismiss T* as hopelessly implausible. But at no point is the bias for T over T* or for V* over V ever supported by evidence. (At least, it is never supported prior to 2000; and it is only prior to 2000 that the bias is important to us.)

even allowing circular 'justifications', since the premisses of an inductive 'justification' of inductive or an inductive-perceptual 'justification' of perceptual rules are empirical. So the issue of the weak apriority of inductive and perceptual rules is largely a verbal issue about the threshold of reasonableness. For the reasons above, I prefer the lower, non-coherentist threshold, on which good inductive and perceptual rules count as weakly a priori. The question of their full apriority then reduces to the question of whether they are empirically indefeasible. I will have something to say in support of the empirical indefeasibility of certain inductive and perceptual rules later, though we'll eventually see that this question too has a quasi-terminological component.

3. DEFAULT *REASONABLENESS* AND THE EVALUATIVIST APPROACH TO APRIORITY

So far the discussion of default reasonableness has been focused more on the 'default' than on the 'reasonableness'. To call a proposition or rule default *reasonable* is to hold that it is *reasonable* to believe or employ it without first adducing evidence or arguments in its favour. Or in other words, that it is *reasonable* to adhere to it as a 'default belief' or 'default rule' (a belief or rule that is accepted or employed without adducing considerations in its favour). The previous section argued (with a slight qualification) that if one is going to have very many beliefs at all one must have default rules; but to get from this to the conclusion that some rules are default *reasonable* and hence weakly a priori, one needs to assume that it is possible to have a sufficient array of reasonable beliefs; and to get to the conclusion that some of the rules *we employ* are default reasonable and hence weakly a priori, one needs to assume that some of our own beliefs are reasonable.

What is it for a default rule (or any other rule) to be reasonable? My main discussion of this will come later, in Section 5, but it will help to give a brief preview now.

One approach to explaining reasonableness (I'll call it 'naturalistic reductionism') has it that the reasonableness of a rule is entirely a matter of how good the rule is at producing truth, avoiding falsehood, and so forth. In the case of deductive rules, we think that ours are objectively correct in that they have complete and non-contingent reliability; and naturalistic reductionism simply identifies this objective correctness with their reasonableness. In the case of inductive and perceptual rules it is less easy to make sense of objective correctness, but we do apparently think that the ones we employ are as a matter of contingent fact reliable, and so are good at arriving at the truth, and naturalistic reductionism simply identifies the reasonableness of the rule with some combination of these and similar 'truth-oriented' characteristics.

In my view, this approach is thoroughly implausible, on numerous grounds. Here is a partial list:

(1) In the case of deductive rules, the notion of reliability is quite clear: and correct rules do have complete and non-contingent reliability while incorrect ones don't. So in this case, the question of whether reliabilism gives the right answer about reasonableness is equivalent to the question of whether it is always reasonable to believe correct logical rules and unreasonable to believe incorrect ones. But I would have thought the answer to be 'no': even if the correct logic for dealing with vagueness or the semantic paradoxes is a non-classical logic (perhaps one that no one has yet formulated), we who do not realize the virtues of such a revision of logic, or even know how to formulate the logic, are not unreasonable in trying to cope with vagueness or the semantic paradoxes in the context of classical logic. We are unreliable but not unreasonable.

(2) The standard 'internalist' criticism: it is implausible to hold that our methods (assuming them reliable in the actual world) would be straightforwardly unreasonable in a 'demon world' (a world designed to make those methods unreliable, but undetectably so).

(3) It isn't easy for a reductionist to satisfactorily explain why a method is unreasonable if it simply builds in an a priori belief in whatever physical theory is in fact correct. (The obvious reductionist approach to explaining that is to require that the method work in certain possible worlds other than our own as well as in our own world; but specifying which other possible worlds are relevant and which aren't, and doing so in a way that isn't grossly *ad hoc*, seems to me extremely difficult.)

(4) The application of the notion of reliability to our basic inductive methods is crucially unclear, for reasons to be given at the end of Section 4; and it is hard to supply a clear replacement for the demand that our basic inductive methods be reliable that isn't either too weak to exclude obviously unreasonable methods or so strong as to further accentuate the problems in (2).

(5) The motivation for reliabilism is suspect: the motivation for wanting our beliefs to be true is clear, and this might motivate an interest in the reliability of a rule as evidence of the truth of beliefs formed by the rule, but it doesn't motivate the stronger role that the reliabilist gives to reliability. More fully: there are lots of classes to which a given belief B belongs such that the proportion of truth to falsehood in that class would have an evidential bearing on the truth of B. If our interest is in the truth of B, we thus have an indirect interest in the proportion of truth to falsehood in many such classes. But the reliabilist, in trying to reduce reasonableness to a precisely defined notion of reliability, wants to single out one particular such class as having a more-than-evidential interest: it's what constitutes the reasonableness of B. Why think that this has any interest?

(6) 'Reliability' is certainly not the only thing we want in an inductive rule: *completely* reliable methods are available, e.g. the method of believing nothing whatever the evidence, or believing only logical truths; but we don't value them, and value instead other methods that are obviously not perfectly reliable, because of their other characteristics. And reliability itself subdivides into many different

notions: for instance, short term vs. long term; yielding a high probability of exact truth vs. yielding a high probability of approximate truth; reliability in the actual world vs. reliability over a range of 'nearby' possible worlds; etc. When one thinks about the different more precise characteristics we value, and the fact that they tend to be in competition with each other, it is hard to imagine how they could be combinable into a package that could plausibly be held to constitute reasonableness.

(7) Familiar worries about naturalistic reductionism in ethics carry over to the epistemological case. For instance, (i) identifying reasonableness with a natural property seems to strip it of its normative force; (ii) in cases of fundamental disagreement about *what* natural property is coextensive with reasonableness, it is difficult to take seriously the idea that one party to the debate is right and the others wrong. (Indeed, that idea seems to presuppose a non-natural property of reasonableness, whose extension is up for grabs.)[14] The naturalist can avoid this by supposing that those in fundamental disagreement about what is reasonable are using the term for *different* natural properties; but this relativist conclusion has the consequence that they aren't really disagreeing, which seems incredible.[15]

Despite all this, I don't think naturalistic reductionism wholly misguided: I hope to provide an attractive picture that captures its insights.

If naturalistic reductionism is rejected, what is to take its place? Another approach is to take our own rules as completely reasonably by fiat, and to regard other people's rules as reasonable to the extent that they are similar to ours. I'll call this 'the egocentric approach'. It too strikes me as hopelessly implausible, this time because it chauvinistically takes our own rules as sacrosanct quite independent of any properties they might have.

What alternatives remain? One could try to combine features of the above approaches, taking reasonableness to be a combination of reliability (and related characteristics) and similarity to ones own methods; but this wouldn't be much better than the egocentric approach as regards chauvinism, and wouldn't help with the main problems with the naturalistic approach either.

[14] This parenthetical point would need to be stated with care, so as not to run afoul of the fact that there are genuinely controversial property-identities (e.g. between being in pain and being in a certain psychofunctional state), and that controversies about them does not necessarily have to be understood in terms of higher-order properties associated with the two terms of the identity, but can be explained in terms of the differing conceptual roles of the terms. But I don't think that the analogy of controversial judgements about reasonableness to controversial judgements about pain does much to raise the plausibility of evaluative naturalism. For one thing, there is a physical property centrally involved in causing our pain judgements, and it seems a fairly straightforward factual question what this is. There seems to be no such straightforward factual question in the case of reasonableness.

[15] Of course, the relativist can admit that the parties disagree in attitude, but in the context of naturalism (or any sort of fully factualist view of reasonableness) this seems *ad hoc*: the natural notion of disagreement for a naturalist (or any sort of factualist) is factual disagreement, and on this notion the parties do not disagree.

A third approach ('non-naturalism') is to regard reasonableness as a primitive property of rules or methods, not explainable in terms of anything non-normative (hence presumably undetectable by ordinary perceptual processes). But what reason would there be to suppose that any rules or methods have this strange property? And even if we assume that *some* have it, what reason would there be to suppose that the rules or methods *we employ* have it? If reasonableness consists in the possession of such a curious property, shouldn't we believe that our rules and methods (and any alternative rules and methods) are unreasonable?

It seems to me that we need another option. The options above had one thing in common: they assumed that reasonableness was a straightforwardly factual property. My proposal is that it is an evaluative property, in a way incompatible with its being straightforwardly factual.[16] We do, I think, evaluate rules and methods in part on the basis of our judgements as to whether using them will be good at leading to true beliefs and avoiding error. We also favour our own method over other quite dissimilar ones. These two strands in our evaluation procedure are inseparable: for (I will argue) we *inevitably* believe that our own methods will be better than the dissimilar methods at leading to truths and avoiding errors. One shouldn't ask whether it is the conduciveness to truth or the similarity to our methods in which the reasonableness consists, for reasonableness doesn't *consist in* anything: it is not a factual property.

The approach I'm recommending ('evaluativism') shares with non-naturalism the conviction that it is quite misguided to try to reduce epistemological properties like reasonableness to other terms. But it is very different from non-naturalism when it comes to the question of scepticism. A sensible evaluativist will think that there are no non-natural properties, or anyway none that is ever instantiated; so that if scepticism were defined as the failure to believe that any rules and methods have such a non-natural property, then the sensible evaluativist is a 'sceptic'. But the evaluativist should say that this is a totally perverse definition of scepticism. On a more reasonable definition, a sceptic is someone who positively evaluates abstention from all belief; scepticism in that sense is idiotic, and surely doesn't follow from the non-instantiation of mysterious properties.

The meta-epistemological views just sketched are important to the interpretation of default reasonableness, and of weak apriority more generally. One kind of

[16] The conception of evaluative properties as 'not fully factual' has been spelled out in different ways. My favourites are Gibbard (1990) and Field (1994). One feature of these views is that they employ a general notion of disagreement that incorporates disagreement in both attitudes and values. When straightforwardly factual matters are at issue, disagreement reduces to factual disagreement. In typical normative disagreement, it is a combination of facts and values that are in dispute. In certain cases of fundamental normative disagreement, no facts are relevant to the disagreement, only values. In this case, the disagreement is of attitudes. But note that this invocation of disagreement in attitudes is very different from the factualist's (n. 15): on a factualist view it is factual disagreement that should be primarily important, so invoking disagreement in attitude seems *ad hoc*; whereas on Gibbard's or my evaluativism, there is only one notion of disagreement, and disagreement in attitude is simply a special case of it.

question about these characteristics is: in virtue of what does a given proposition or method have them? In virtue of what is it reasonable to use modus ponens on no evidence?[17] The difficulty of providing an answer to this question is one of the main reasons that apriority has seemed mysterious. The meta-epistemology I've suggested requires that this question be recast: the proper question is, why value a methodology that allows the use of modus ponens on no evidence? Well, one needs some methodology, so the question can only be why favour this methodology over alternatives, and the answer will depend on what alternative methodologies are possible. The alternatives to a methodology that allows use of modus ponens on no evidence divide between those that license its use on certain empirical evidence (maybe on the evidence that snow is white?) and those that don't license its use at all (but license no deductive inference at all, or license only some particular weak logic that doesn't include it). The question then reduces to showing what is wrong with particular methodologies of each type. I don't want to get into a substantive discussion of what is wrong with particular methodologies of each of these types; my point is only that that is what is involved in defending the weak apriority of modus ponens, once one adopts the evaluativist perspective. This seems to me a substantially different perspective on a priority than one gets from more fully 'factualist' meta-epistemologies, and this different perspective does a great deal to remove the mystery from weak apriority.

It isn't just issues about weak apriority that evaluativism recasts; issues about empirical indefeasibility are recast as well. For an evaluativist, defending the empirical indefeasibility of modus ponens is a matter of arguing that a methodology that takes it as empirically indefeasible is preferable to methodologies that allow it to be empirically defeated by particular kinds of evidence. If an anti-apriorist charges that it would be dogmatic for a system of rules to license the use of modus ponens *on any evidence whatever*, the response should be 'This is better than their licensing its revision on inappropriate evidence (say, the discovery of a new kind of virus); give me a plausible example of possible evidence that would make it *appropriate* to abandon modus ponens! And if the possible evidence you cite isn't *obviously* appropriate for this purpose, then give me at least a sketch of a theory of evidence on which the evidence *is* appropriate!' Without even a sketch of an answer, it is hard to see why we should take empiricism about modus ponens seriously. I don't say that we ought to rule out the possibility that someone could come up with an example of possible evidence that would seem to make it appropriate to give up modus ponens (were the evidence actual), and of a possible theory of evidence that explained why this was evidence against the adequacy of modus ponens. But ruling out that possibility is something no a

priorist should try to do: however apriorist we are about logic, we ought to be fallibilist enough to recognize the possibility that new conceptual developments will undermine our apriorism.[18]

Incidentally, the failure to distinguish apriorism from infallibism about apriorism seems to underlie the widespread belief that Quine has provided an alternative to apriorism about logic. Quine's view is that one should evaluate alternative logics in combination with theories of the rest of the world: given a theory of everything, including a logic, one uses the logic in the theory to generate the theory's consequences. Then we choose a theory, including a logic, on such grounds as overall simplicity and conservativeness and agreement with observations. But this description of the methodology is so vague that it is not in the least clear that it dictates that modus ponens or any other part of logic should be revisable on empirical evidence. Whether it so dictates depends on the standards of simplicity and conservativeness of overall theories: it depends on whether the decrease of simplicity and conservativeness that results from modifying the logic could be compensated by incerased simplicity and conservativeness in other parts of the theory (holding the observational predictions fixed). It is *conveivable* that the standards of simplicity we use, or attractive alternative standards, will be such as to say that there are possible observations that would lead to favouring a theory that includes an alternative to our logic over ours. That is enough to undermine infallibilism about apriority, but to undermine a priority one must show that there actually are attractive standards of simplicity under which possible observations would lead to an alternative logic, and Quine has given no clue as to what those standards of simplicity might be. (Indeed there is reason for scepticism about the existence of standards that would let possible observations undermine modus ponens. For one is likely to need modus ponens in the background logic in which one reasons about what follows from each theory-plus-logic and how well it accords with observations; and it is hard to imagine that a person using this background logic could rationally come to favour a theory-plus-logic in which the logic conflicted with the background logic.)

I don't pretend that this discussion settles the case for a priorism about logic: it is intended only to illustrate how evaluativism supplies a perspective for answering the question that does not turn on rational insight into the nature of non-natural epistemological properties.[19]

[17] If default reasonableness rather than weak a priority is in question, I should say 'on no evidence *or argument*'. But presumably we attach little importance to the difference between a methodology that takes modus ponens to be default reasonable and one that takes it to be weakly a priori because derivable from disjunctive syllogism which is in turn taken as default reasonable.

[18] More on this and some of the other claims in this paragraph and the next is to be found in Sections 2 and 4 of Field (1998*b*).

[19] One of the important issues not addressed is whether Benacerraf's (1973) puzzle about how a priori mathematical knowledge is possible extends to other alleged cases of a priori knowledge. I think Benacerraf's argument does work against many claims to apriority. (Including claims of a priori access to mathematical entities *as these are conceived by most Platonists*. For a discussion of which Platonist views might survive the argument, see Field (1998*a*).) For instance, the reasons for negatively evaluating a system of rules that would allow us to adhere *whatever the evidence* to a particular physical theory that we hold true have to do with the fact

4. AN EPISTEMOLOGICAL PUZZLE

I turn next to the question of whether our inductive and perceptual methodologies are best viewed as empirically indefeasible; this will lead into further discussion of reliabilism and evaluativism. A good way into these issues is by way of an epistemological puzzle.[20] It comes in two parts.

Part One starts with the idea that we want our empirical methods to be reliable: to lead to a fairly high proportion of true beliefs. In particular, we want them to be reliable in present and future applications. But we can empirically investigate whether they have been reliable in past applications; and it is surely possible that we will discover that one of our empirical methods hasn't been reliable, and that some alternative to it has been. Alternatively, we might discover that our method has been *fairly* reliable, but that some alternative method has done much better. If we did discover one of these things, then since we take the past to be the best guide to the future we should conclude that our method will continue to be less reliable than the alternative. But surely if we think that one of our own empirical methods will be less reliable than an alternative, then we ought to switch from our method to the other. All of this would seem to apply not only to our 'non-basic methods'—our rules of thumb (like 'Believe what the *NY Times* says') that are evaluated using more basic methods (like induction); it would seem to apply to our most basic inductive method itself. That is, it would seem that we want our most basic inductive method to be reliable, and can investigate empirically whether it has been, and we will stop using it if we find that it has not. But in this case, the investigation of the most basic method can't be by another method, for by hypothesis none is more basic. Rather, the investigation of our most basic method uses that very method. So in the case where we empirically discover that method unreliable, the decision not to use the method would be based on that very method.[21]

Part Two of the puzzle says that the conclusion of Part One is incoherent. How can our method, in combination with evidence E (in this case, evidence of its own unreliability), tell us not to follow that very method? Our method presumably already tells us something about what is legitimate to belief and what is illegitimate to believe when our evidence includes E (say, when it consists of E&F).[22] These instructions might be empty: they might allow us to believe what we like. Or they might tell us to stop believing. Or they might tell us any of numerous other things. But whatever they tell us it's legitimate to believe on E&F, that's what we must do if we are to follow the method. Now if the method tells me that E undermines the method, it must tell me not to always do what the method tells me to do; in other words, it must tell me to do something different, on some evidence E&F, from what it tells me to do on E&F. In other words it must offer me inconsistent instructions. It would seem that only an inconsistent empirical method can allow itself to be undermined by empirical evidence of its own past unreliability; to suppose that good empirical methods must allow themselves to be empirically undermined in this way is then to suppose that good methods must be inconsistent, which seems absurd.

To summarize: Part Two is an argument that we can't possibly treat our basic empirical methods as empirically defeasible, whereas Part One is an argument that we must do so; where to take a method *as* empirically defeasible is to adopt standards on which empirical evidence could count against it. Obviously something is wrong, but what?

A superficially plausible diagnosis is that the key error was the presupposition that there is such a thing as 'our basic empirical method': that is, in the supposition that we employ a method that can't be undermined by any of our other methods. One might argue that this supposition is incorrect, that in fact we employ many different methods, each of which can be assessed using the others. I think myself that the assumption of a basic inductive method is defensible if properly understood, but I will save that issue for an appendix. What I want to do now is argue that the issue of whether there is a basic method isn't central to the puzzle, because there is a more straightforward error in the argument in Part One.

In investigating this, it helps to be a bit more concrete: instead of talking about 'our basic empirical method' let's instead talk about a specific inductive rule. For simplicity I'll pick an extremely simple-minded rule, but one which I think will have crucial features in common with any inductive rule that might plausibly be regarded as part of our basic inductive method. The rule I'll pick for illustration is the following:

that doing so would clearly make our belief causally and counterfactually independent of the facts; and such independence from the facts would defeat the epistemological value of the considerations on which the belief was based. (I think that is what a Benacerrafian argument against apriority about physics would amount to.) It might seem that this would apply equally well to apriority about logic. The idea would be that a priori belief in logic makes logical beliefs similarly independent of the facts, and that this is equally bad. But I think that in the logical case one simply can't make sense of the question of whether logical beliefs depend on the logical facts; so we can't make sense of the claim that is supposed to defeat the evidential value of the considerations on which the belief was based, and so the logical beliefs remain undefeated. For more details, see Field (1996: sect. V) or Field (1998*b*: sect. 5).

[20] The puzzle is implicit in many epistemological writings; probably its most explicit presentation is as the argument against 'norm externalism' in Pollock (1987) (though it is close to explicit in Putnam (1963) and Lewis (1971).) My resolution is not too far from Pollock's, though Pollock's view is closer to what I've called the egocentric approach than to the evaluativism I recommend.

[21] It has been suggested to me in conversation that our basic method is the meta-method 'employ whatever first-order method is most reliable'; and that this meta-method couldn't fail to be reliable. But in fact the proposed meta-method is not an employable method. To make it into one, we would need to recast it as something like 'employ whatever first-order method you believe to be most reliable', or 'employ whatever first-order method your first-order methods tell you to believe most reliable'; and these *can* certainly fail to be reliable.

[22] Our method may take into account factors other than the available evidence—for instance, it may take account of which theories have been thought of—but as far as I can see, such additional factors won't matter to the argument that follows.

(R) If in the past you have observed n ravens, and m of them have been black, you should believe to degree $(m+j)/(n+k)$ of any raven not yet observed that it is black,

where j and k are fixed real numbers, $0<j<k$.[23] The idea of the rule is that j/k is your initial degree of belief that a raven will be black; as observations of raven colour accumulate, that initial degree of belief gradually becomes swamped by the observed frequency. (This happens slowly if k is large, quickly if k is small.) This is of course a thoroughly implausible rule: among other defects, (i) it takes no account of any evidence other than observations of raven colour; (ii) it takes no account of any regularities in the *ordering* of black and non-black among the ravens observed; and (iii) it takes no account of how the ravens were selected for observation. When I say that the rule is implausible, part of what I mean is that our own basic inductive rule does not have these limitations. But let's pretend that we do employ this rule, and let's ask how if at all empirical evidence (e.g. of the past unreliability of the rule) could lead us to rationally revise it.

Since the rule is one for degrees of belief rather than for all-or-nothing belief, talk of reliability or unreliability may not be strictly appropriate; but clearly the analog of a discovery of past unreliability in the rule is the discovery that the actual proportion of blackness among ravens observed in the past has been substantially different from j/k. The argument of Part One suggests that were we to discover this, then those observations would provide evidence against the rule. But this is mistaken. There is of course no doubt that if j/k is small, then the observation of many ravens with a high proportion of blackness among them should lead us to revise the probability of blackness for an unobserved raven upwards (barring special additional information anyway). But this doesn't mean that we should modify the rule: our observation of a high proportion of blackness among ravens is something that the rule takes into account. Suppose our initial bias was $j = 1$, $k = 10$, so that the degree of belief assigned to a given raven being black was only 0.1. And suppose we observe 20 ravens, of which 19 are black. Then the rule tells us to believe to degree 0.667 (20/30) of an unobserved raven that it is black. The rule has in a sense told us to modify our biases. In another sense, though, the biases go unchanged: the initial bias, represented by the pair <j,k> from which we started, is still there producing the new degree of belief. Of course, the initial bias produces that new degree of belief only with the accumulated evidence, and *the effect of using the initial bias with the accumulated evidence is in a sense equivalent to altering the bias*: it is equivalent to altering the initial bias to $j = 21$ and $k = 25$, *and then dropping the observation of the first 20 ravens from our evidence*. (It would be double counting to let it alter the bias and then *in addition* let it count as evidence with the new bias.) In a sense then our rule is 'self-correcting': with accumulating evidence the old rule behaves like an altered rule would *without* the added evidence. Because it is 'self-correcting' in this way, there is no need for it to be revised when evidence of its unreliability in previous applications is accumulated.

Of course there *are* ways that a rule like (R) might be inductively challenged. (R) was supposed to be a rule for ravens; and if we had a great deal of experience with other birds showing that a high degree of concentration in colour tends to prevail within each species, that would give us grounds for lowering the correction factor (lowering the j and the k while keeping their ratio constant) in the raven rule. But that simply shows that the original rule isn't a serious candidate for a basic inductive rule. Any serious candidate for a basic inductive rule will allow 'cross-inductions': it will allow for evidence about other birds to affect our conclusions about ravens.[24] Were we not to employ a rule that allows cross-inductions, we wouldn't regard the evidence about other species as relevant to ravens, and so would not see such evidence as providing any reason to lower the correction factor in (R).

So the point is that we use a more complicated inductive rule than (R) (one that 'self-corrects' to a greater extent than R does); using the more complicated rule, we can inductively conclude that our future practice should not accord with the more simple-minded rule. But if we had used the more simple-minded rule in doing the assessment, we wouldn't be able to conclude that our practice should no longer accord with that simple-minded rule; similarly, I suggest, if we tried to assess the more complicated rule using the more complicated rule, we couldn't ever recognize anything as evidence for the conclusion that we shouldn't use it. We could recognize evidence that our rule hadn't worked well in the past, but this would simply be evidence to be fed into the rule that would affect its future applications; we would not regard it as evidence undermining the rule. (Some apparent objections to this are discussed in the appendix.)

What I have said suggests that if indeed some inductive rule is basic for us, in the sense that we never assess it using any rules other than itself, then it must be one that we treat as empirically indefeasible (hence as fully a priori, given that it will surely have default status). So in the puzzle, the error in reasoning must have come in Part One. Moreover, it is now clear just where in Part One the error was: the error was in supposing that because the rule had been unreliable in the past it was likely to be unreliable in the future. What the discussion shows is that *there is no reason to extrapolate from the past in this way: for the future employment of the rule takes account of the unreliability of the past employments, in a way that makes the future applications qualitatively different from the past employments*. That's so in the case of rule (R), and it will also be so in the case of any

[23] The rule is to be generalized to apply to other pairs of predicates besides 'raven' and 'black', though as the 'grue' paradox makes vivid, using the rule for some pairs requires not using it for certain others. The rule given is in effect an instance of Carnap's λ-continuum (at least when k/j is an integer ≥2): it results from taking λ=k and the number of kinds as k/j.

[24] Kemeny and Carnap investigated how to expand Carnap's λ-continuum to include a new parameter that allows for such cross-inductions: see Kemeny (1963: 732-3) and Carnap (1963: 977). The resulting formulation of an inductive method is still quite simplistic, but a step in the right direction.

sufficiently general inductive rule that anyone would be tempted to employ. So to resolve the puzzle there is no need to deny the existence of basic rules, we must merely recognize that any plausible candidate for a person's basic rule will have the same 'self-correcting' character as (R).[25]

I have been discussing whether we could empirically overturn a most basic inductive rule, and concluded that we could not. If so, this also calls into question the idea that we could ever rationally accept anything as empirically *supporting* these rules: for it is hard to see how there could be possible observations that support the rules without possible alternative observations that undermine them. The idea that our most basic inductive rules could get inductive support has been much discussed in the context of the justification of induction. It has often been dismissed on the ground that an inductive justification of induction is circular, though as noted earlier a number of authors have argued that the sort of circularity involved here (rule-circularity) is not vicious. I agree with these authors that the kinds of arguments that are offered as rule-circular justifications are of interest; but their worth *as justifications* turns on the idea that they remove a prima-facie risk, a risk that reasoning with the rule will lead to the rule's invalidation. In the case of inductive justifications of induction, what they turn on is the idea that the basic inductive rule might be inductively undermined; and that, I am suggesting, is impossible. It is that, not rule-circularity in itself, that is the real reason why the inductive justification of induction is an illusion. (Something similar holds of deductive justifications of deduction, I believe.)

Before drawing further morals, it will be helpful to consider another illustration, this time involving a perceptual rule. It is natural to suppose that rules of perception can be empirically overturned. Suppose we are initially disposed to regard things that look a certain way 'red'. (I'll pretend here that how things look to us is independent of our tendencies to judge their colour: it won't affect the point.) We then discover that in certain lighting, things that look that way often aren't red; and using this information, we revise our practice of making colour judgements. So it looks like our initial 'rule of perceptual judgement'

> (P) Believe things red if they look red

has been overturned. But that, I submit, is misleading: the right thing to say rather is that our initial practice was sensitive to inductive considerations that weren't built into (P), so that (P) isn't the most basic rule we were following (even before the information about the lighting). After all, if it had been the most basic rule we were following, it is hard to see how the information about the lighting could have influenced us.

[25] Less inductively sophisticated creatures doubtless employ simpler inductive methods that are not 'self-correcting' in this way. Such creatures could either never have evidence for the past unreliability of their methods, or could never think to extrapolate it, or would continue reasoning as before despite the belief that their methods would be unreliable. But they aren't us.

One way to think about what's going on in this case concedes that we do employ (P), but only as a default rule. The more basic rule is a meta-rule that says: use (P) unless it is inductively shown unreliable in certain circumstances; if that happens, stop using it in those circumstances. The meta-rule allows (P) to be inductively overturned, but it's hard to see how the meta-rule itself can be inductively overturned: we treat the meta-rule as empirically indefeasible (indeed, as a priori).

We don't really need the 'default rule'—'meta-rule' contrast, we can build everything into the ground level rule, by taking that rule not to be (P) but rather something more like

> (P*) Believe a thing red if it looks red, unless there are circumstances C that you believe you're in such that you have inductive evidence that looking red is misleading in those circumstances

or

> (P**) The extent to which you should believe a thing red should be such and such a function of (i) how it looks to you; (ii) your evidence about the perceptual circumstances; (iii) your background evidence about how the way that things look depends on the perceptual circumstances; and (iv) your background evidence about what colour it actually is.

Here the point is that inductive evidence that the rule has failed in the past feeds into the rule, in a way that alters the results of applying the rule in the future: for instance, in (P**) the evidence does this by affecting our beliefs of type (iii), on which the degree of belief that the thing is red depends. We have the same situation as with the inductive rule (R): the relevance of evidence of the past unreliability of the rule isn't to undermine the rule; rather, the evidence is something *that the rule will itself take account of, and that will substantially modify the future applications of the rule (in a way that might be expected to make those future applications more reliable than the past ones)*. Of course, not every conceivable rule will 'self-correct' (in this sense) on the basis of evidence of its past unreliability; but those rules that we in fact take seriously do.[26] Such general rules are never really put to risk in inductive assessment; all that is put to risk is the particular manner in which they've been employed. And again, that means that the idea

[26] So even if we had been unlucky enough, or evolutionary maladapted enough, to employ rules which gave initial weight to our purported telepathic experiences in addition to our perceptual experiences, then as long as those rules were analogous to (P*) or (P**) rather than the cruder (P), we would have long since discounted telepathy. (This is in response to Goldman (1980: 42); although Goldman's formulation, and the surrounding discussion, seems to depend on (i) his temporarily assuming that we *choose* our basic inductive and perceptual rules, and (ii) his assuming that what we are after is finding *the uniquely correct* epistemological rules. I want no part of either assumption.)

of inductive justification of the rules, which requires the rules themselves to be put to risk but to survive the challenge, doesn't get off the ground.

The bearing of all this on the apriority (or empirical indefeasibility) of inductive and perceptual rules is not entirely direct. For one thing, the discussion has been premissed on the supposition that some inductive rule is basic for us. Whether that is so is a matter I discuss in the appendix: I argue that the question has a quasi-terminological component, but that there are considerations that favour a positive answer. But even given a positive answer to this, what I have argued hasn't been that our most basic rules are a priori or empirically indefeasible; it has been that we *treat them as* empirically indefeasible and indeed a priori: we *don't regard* anything as evidence against them. For a non-evaluativist, this distinction is crucial. For instance, a non-naturalist will say that the non-natural evidence relation may well hold between some evidence and an alternative to one of our basic rules, even though we could never be persuaded to adopt such an alternative on the basis of that evidence by principles of evidence we accept; and a reliabilist will say something analogous. From an evaluativist perspective, though, the distinction is rather academic: the only question is whether we should accept any possible evidence as undermining our rule, and since the rule itself is employed in making the evaluation of what we should do, there is no chance of a positive answer. More on this perspective in the next section.

The examples of (R) and (P*) or (P**) have an importance beyond their bearing on the empirical indefeasibility of our empirical methodology: they also create a problem of interpretation for many versions of naturalistic reductionism. According to naturalistic reductionism, the reasonableness of an epistemological rule *consists in* its having a certain combination of truth-oriented properties, and most advocates of naturalistic reductionism place 'reliability' high on the list of the properties that a reasonable rule must have. But as applied to 'self-correcting' rules like (R) and (P*) or (P**), it is not entirely clear what 'reliability' comes to (even if the reliability of a rule is assessed relative to the specific circumstances of application within a possible world, rather than assessed in the possible world as a whole).[27]

[27] Goldman (1988) offers a substantial reason for *not* relativizing to circumstances within a possible world in assessing reliability: if one is allowed to do so, what is to keep one from so narrowing the circumstances that they apply only to one instance? (That would mean that whenever the rule yields a truth, however accidentally, it would come out reliable, and so following it on that occasion would count as reasonable according to reliabilist lights.) Perhaps there are ways to block carving the circumstances so finely, but it isn't in the least clear how to do so without gross ad-hocness; and so Goldman adopts for not allowing any relativization to circumstances within a world. (Indeed, he argues that even consideration of reliability within a possible world as a whole is too narrow: one must consider reliability with respect to a class of similar worlds. The motivation for doing this is so that a rule that 'happens' to yield truths about a particular world independent of evidence won't count as reliable. At this point one might raise the question of how to carve out the relevant class of worlds without gross ad-hocness, but I will not press the matter.)

We can see this in the case of (R) by imagining that there was a fairly strong initial bias (moderately large k), and the initial degree of belief j/k differed drastically from the actual frequency of blackness among ravens: perhaps j/k is quite small whereas the proportion of blackness among ravens is very high. (For simplicity I will confine my attention to the case where that proportion is fairly stable from one epoch to another and one region to another.) And suppose that (R) is applied long enough for the initial bias to become largely swamped by the observations. On the question of whether the use of the rule counts as reliable, there seem to be three options:

(i) We can say that the rule was not reliable in its early uses (prior to the swamping), but became so later (after the swamping); after all, the degrees of belief produced by late uses of the rule closely reflected the actual frequencies, but degrees of belief produced by early uses were wildly at variance with actual frequencies. (Of course the swamping is gradual, so the shift from unreliability to reliability is gradual.)

(ii) We can say that the rule was not reliable in early or late uses: the fact that it initially produces degrees of belief wildly at odds with frequencies shows that it simply isn't a reliable rule, but merely gives results in its later uses that closely match the results given by reliable rules (those with a more optimal ratio j/k).

(iii) We could say that it was reliable in both: that the apparent unreliability in early uses results from taking too short-term a perspective.

Which option should a reliabilist adopt? Given reliabilism, (ii) would make reasonableness hard to come by: a faulty bias would doom us to unreasonableness forever (barring a shift in inductive policy that is not evidence-driven). I think that (i) accords best with the normal use of 'reliable'. However, given reliabilism, (i) requires that the use of the rule was unreasonable at first but became reasonable as the rule was used more and more; this strikes me as somewhat counter-intuitive, and it is contrary to the doctrines of at least one prominent reliabilist: see note 27. Some reliabilists might then prefer the long-run perspective implicit in (iii): even early uses of the rule count as reliable, because the rule would yield good results if applied in the long run. If 'long run' here means *really* long run, this would be even more counter-intuitive than (i): 'dogmatic' versions of (R) with exceptionally large k that would take millions of observations to significantly swamp would count as reliable and hence reasonable. It would also blunt the force of reliabilism, in that very few rules would be declared unreliable. But a reliabilist could avoid this by adopting (iii) for the case under discussion, where k is moderately large, and adopting view (i) or (ii) in the case of exceptionally large k where it would take a very long time for swamping to occur; in effect this is to use an intermediate length of reliability as a criterion of reasonableness for the early uses of the rule. This combination gives the most intuitively appealing results about reasonableness. But it is not clear that this is in keeping with the spirit of reliabilism: for it is now a priori (relative anyway to the assumption of stability in proportions) that *all* versions of (R) where k is not too large are reliable from the start, whatever the value of j (greater than 0 and less than k); the idea that the facts about the actual world determine an inductive method uniquely or close to uniquely (see Goldman 1980) is completely abandoned.

Something similar holds for perceptual rules like (P*) or (P**). Imagine a world where deceptive appearances are common enough so that in the initial stages of use the rule leads to error a quite substantial per cent of the time, but not so common as to prevent the rule from ultimately 'self-correcting' if appropriate observations are made. We can again ask whether uses of the rule are 'reliable' and hence reasonable before such a 'self-correction' and whether they are reliable and hence reasonable afterwards. (Actually there are two important differences between this case and the induction case: first, it is likely to take much longer for the rule to self-correct; second, the self-correction is not automatic, in that whether a self-correction is ever made is likely to depend on accidents of which observations are made and which theories are thought of. I think that both of these factors diminish the chance that a reasonable long-term perspective on reliability could rule the early uses of the rule 'reasonable'.) In this case too it is unclear what a reliabilist can say that keeps the spirit of reliabilism without making reasonableness implausibly hard to come by.

5. MORE ON EPISTEMOLOGICAL EVALUATIVISM

I propose an alternative to reliabilism, more in line with 'non-factualist' views about normative discourse. The alternative is that reasonableness doesn't consist in reliability or anything else: it's not a 'factual property'. In calling a rule reasonable we are evaluating it, and all that makes sense to ask about is what we value. So the relevance of the reliability of a rule to its reasonableness is simply that we place a high value on our inductive and perceptual rules leading to truth in the circumstances in which we apply them; more or less equivalently, we place a high value on a given rule to the extent that we *believe* it will be reliable in the circumstances in which we apply it. We saw earlier that one will inevitably believe our most basic rules to be reliable in the circumstances in which we intend to apply them.[28] If so, we will inevitably place a high value on our own inductive and perceptual rules.

Is this an 'externalist' view or an 'internalist' view? The answer is that that distinction as normally drawn (for instance in Goldman (1980)) rests on a false presupposition. The presupposition is that epistemological properties like reasonableness are factual. If they are factual, it makes sense to ask whether the factual property involved includes 'external' elements.[29] On an evaluativist view, it is hard to draw a distinction between externalism and internalism that doesn't collapse. Any sensible evaluativist view will be 'externalist' in that one of the things we value in our rules is (some restricted version of) reliability. A sensible view will also be 'internalist' in that we also place a high value on our own rules: indeed, those are the rules we will use in determining the reliability of any rules we are evaluating. Which is primary, the high valuation of *our own* rules or the high valuation of *reliable* rules? It is hard to give clear content to this question, since (by the previous section) we inevitably ought to regard our own rules as likely to be reliable in the circumstances in which we intend to apply them.[30]

A view like this raises the spectre of extreme relativism. For mightn't it be the case that different people have different basic standards of evaluation? If so, aren't I saying that there is no fact of the matter as to which standard of evaluation is correct? And doesn't that mean that no standard is better than any other, so that those who 'justify' their belief in reincarnation on the basis of standards that positively evaluate just those beliefs that they think will make them feel good about their cultural origins are no worse epistemologically than those with a more 'scientific bias'? That of course would be a totally unacceptable conclusion.

But nothing I have said implies that no standards are better than others. Indeed, some clearly are better: they lead both to more truth and to less falsehood. Of course, in saying that that makes them 'better' I am presupposing a goal that is being rejected by the imaginary 'feel gooders', but so what? All evaluations presuppose goals, and of course it is my own goals that I presuppose in making evaluations. (To paraphrase David Lewis: Better I should use someone else's goals?)

Not only must I presuppose my own goals in saying that my standards are better than others, I must presuppose my own beliefs. This is most easily seen by contrasting my own standards not to 'feel good' standards but to the standards of those who are interested in the truth but have bizarre views about what best achieves it (e.g. they accept counter-inductive methods, or believe whatever the Reverend Moon tells them). If one were to apply the methods such people accept, one would be led to the conclusion that their methods are better than scientific methods. But again, so what? What makes scientific methods better isn't that *they say that* they will lead to more truth and less falsehood than these other methods.

[28] More accurately, our rules license us to so believe.

[29] I assume for present purposes that the contrast between external and internal elements is clear.

[30] One could hope to make sense of it by considering conditional evaluations: we ask people to evaluate certain rules *on the supposition that* they are reliable, or unreliable. For instance, we consider the possibility of a world where our methods are unreliable and methods we find bizarre are reliable, and ask whether our method or the bizarre method is 'reasonable' in that world. But it seems to me that when asked this question we are torn: the two strands in our evaluation procedure come apart, and what to say is simply a matter of uninteresting verbal legislation.

it is that *they do* lead to more truth and less falsehood than these other methods. In saying that they do this I am presupposing the methods I accept, but that should go without saying: that's what accepting a method involves.

Of course, this is circular. ('Rule-circular', anyway). Earlier I objected that in using a methodology to evaluate itself, a positive evaluation isn't worth much *as a justification* of the methodology unless there was a prima-facie risk that the evaluation would have turned out negative; and that with our most basic rules there is no such risk. But I conceded that rule-circular 'justifications' of our methods have another role: they serve to explain why we value our methods over competing ones. It is that point I am stressing here, and it is enough for the point at hand; for the point at hand was that it is not part of the evaluativist view in question that all methods are equally good. (I'm not now addressing the sceptical issue: to what extent are we *reasonable* in thinking that our methods are better than others. I'll address that soon.)

Returning to the 'argument' for extreme relativism, I think we should concede that different people have slightly different basic epistemological standards: for one thing, any serious attempt to formalize inductive methods always invokes a number of variable parameters ('caution parameters' and the like), and there seems no motivation whatever for supposing that these are the same for all people. I doubt that there are many people with *radically* different basic epistemological standards, though there may be some: in the case of the Moonies it is hard to know what epistemological standards might have been used in producing their beliefs. But the extent of variation is a sociological issue on which I do not want to rest my views: whatever the extent of the actual variation in basic epistemological standards, there might have been such variation—even radical variation. Given that there is possible variation in basic standards (whether moderate or radical), should we suppose that some standards are *correct* and others *incorrect*? I doubt that any clear sense could be given to the notion of 'correctness' here. If there were a justificatory fluid that squirts from evidence to conclusions, we could say that correct standards were those that licensed beliefs in proportion to the fluid they receive from available evidence; but absent that, it is hard to see what can make standards correct or incorrect. What we *can* say is that some standards are better than others in achieving certain goals; and to the extent that one adopts those goals, one can simply say that some standards are better than others. Even given the goals, talk of 'correct' standards is probably inappropriate: for if it means 'best' there may be no best (there could be incomparabilities or ties; and for each there could be a better); and if it means 'sufficiently good', then it blurs relevant differences (two methods over the threshold would count as both correct even if one were better than the other).[31] We need a goal-relative notion of better

standards, not a notion of correct standards. The argument for extreme relativism failed primarily in the slide from 'there are no correct standards' to 'all standards are equally good.'

The position I'm advocating does allow for a sort of moderate relativism. For in evaluating systems of epistemological rules, we can recognize that certain small modifications would produce results which have certain advantages (as well as certain disadvantages) over the results ours produce. For instance, we recognize that a system slightly more stringent in its requirements for belief is more reliable but less powerful. So we recognize that a slight modification of our goals—an increase in the relative value of reliability over power—would lead to a preference for the other system, and we regard the alternative goals as well within the bounds of acceptability. Consequently we make no very strong claims for the preferability of our system over the alternative: the alternative is slightly less good than ours given our precise goals, but slightly better on alternative goals that are by no means beyond the pale. 'Relativism' in this weak sense seems to me an eminently attractive position.

(Pollock (1987: sect. 4) tries to avoid even this weak form of relativism, by proposing that each person's concepts are so shaped by the system of epistemological rules that he or she employs that there can be no genuine conflict between the beliefs of people with different such systems; as a result, the systems themselves cannot be regarded as in conflict in any interesting sense. But this view is wholly implausible. I grant that there's a sense in which someone with even slightly different inductive rules inevitably has a slightly different concept of *raven* than I have, but it is not a sense that licenses us to say that his belief 'The next raven will be black' doesn't conflict with my belief 'The next raven will not be black.' It seems hard to deny that there would be a conflict between these raven beliefs, and if so, the systems of rules give genuinely conflicting instructions.)[32]

A complaint about evaluativism that has sometimes been made to me in conversation is that it places no constraints on what one's epistemological goals ought to be: nothing makes it *wrong* for a person not to care about achieving truth and avoiding falsehood, but care only about adopting beliefs that will make him feel good about his cultural origins. But I'm not sure what sort of ought (or what sort of wrongness) is supposed to be involved. If it's a moral ought that's at issue,

[31] One way to see the importance of this is to suppose that standards improve over time, and that a certain belief B counts as reasonable on the evidence E available at t using the quite good standards S in use at t, but counts as unreasonable on the same evidence using slightly better standards S′ that only become available later (but which might in turn, for all we know, eventually be superceeded). Any attempt to describe this situation in the language of 'correct standards' loses something important.

[32] Pollock's view is that it is our object level concepts like *raven* that are determined by our system of rules. A slightly more plausible view is that our epistemological concepts like *reasonable* are so determined: 'reasonable' just means 'reasonable according to our (the assessor's) rules'. But that view wouldn't serve Pollock's purposes: the advocates of alternative systems of rules would still be in genuine conflict about ravens, and each could raise sceptical worries about whether it mightn't be better to shift from the system that is reasonable in their own sense (viz. their own system) to the system that is reasonable in the other person's sense (viz. the other's system).

fine: I'm not opposing moral standards on which one ought to aim for the truth. But I assume that what was intended was not a moral ought, but some sort of epistemological ought. And that gives rise to a perplexity: on the usual understanding of 'epistemological oughts' they govern beliefs, not goals, and I have no idea what the sort of epistemological ought that governs goals could amount to.

As for 'constraints' on epistemological goals, again I don't think that the intended sense of 'constraint' is intelligible. If McRoss's main goal in forming beliefs is making himself feel good about his cultural origins, well, I don't approve, and I might try to browbeat him out of it if I thought I could and thought it worth the trouble. If I thought that my telling him he OUGHT not have such goals would influence him, I'd tell him that. Is this saying there are 'constraints' on his goals? Nothing is constraining him unless he takes my social pressure as a constraint. But if the question is whether there are constraints in some metaphysical sense, I don't think the metaphysical sense is intelligible. We don't need to believe in metaphysical constraints to believe that he's got lousy goals. (And if calling the goals lousy is evaluative rather than factual, so what?)

Perhaps talk of 'metaphysical constraints' on goals is supposed to mean only that McRoss's goals shouldn't count as 'epistemological'. Or alternatively, that the so-called 'beliefs' arrived at by a system motivated by the satisfaction of such goals shouldn't count as genuine beliefs. I have nothing against 'metaphysical constraints' in one of these senses, though they might better be called 'semantic constraints': they are simply stipulations about the meaning of 'epistemological goal' or 'belief', and of course one may stipulate as one likes. Such stipulations do nothing to constrain McRoss in any interesting way: if he has goals that don't satisfy my constraints, why should he care whether I call his goals 'epistemological' or his mental states 'beliefs'? Nor is it clear what other useful purpose such stipulations might serve.

As I've said, I doubt that there are many people with such radically different epistemological (or schmepistemological) goals for forming beliefs (or schmeliefs). But their non-existence has nothing to do with 'metaphysical constraints': as Richard Jeffrey once remarked, 'The fact that it is legal to wear chain mail on city buses has not filled them with clanking multitudes' (Jeffrey 1983: 145).

Let's now turn to a different complaint about evaluativism: this time not about the lack of objectivity in the goals, but about the lack of objectivity in the beliefs even when the goals are fixed. One way to press the complaint is to make an unfavourable contrast between evaluativism in epistemology and evaluativism in moral theory. In the moral case, an evaluativist might stress the possibility of variation in moral goals (e.g. with regard to the respective weights given to human pleasure and animal pain), but agree that relative to a choice of moral goals some moral codes are objectively better than others, and that we can make useful evaluations as to which ones are better given the goals. Such evaluations are in no way circular: in evaluating how well a given moral code satisfies certain goals, one may need to employ factual beliefs (for instance, about the extent of animal suffering), but such factual beliefs can be arrived at without use of a moral code. In the epistemological case, however, the evaluation has the sort of circularity that has cropped up several times already in this paper: in assessing how well a system of inductive or perceptual rules satisfies goals such as reliability, one needs to use factual beliefs, which in turn are arrived at only using inductive or perceptual rules. And this circularity might be thought to somehow undermine evaluativism—either directly, or by leading to a sceptical conclusion that makes the evaluativism pointless.

The circularity is undeniable: it might be called the fundamental fact of epistemological life, and was the basis for the puzzle in Section 4. But it doesn't directly undermine evaluativism, it leads only to the conclusion that our basic system of inductive rules (if indeed we have a basic system) is in Lewis's phrase 'immodest': it positively evaluates itself over its competitors (Lewis 1971: 1974). Nor is scepticism the outcome: true, systems of rules that we don't accept lead to different evaluations than ours do, but why should that undermine the evaluations provided by the rules that we do accept?

I concede that in dealing with people who use different standards of evaluation from ours, we typically don't just insist on our standards: we have several techniques of negotiation, the most important of which is to evaluate on neutral grounds. And to some extent we can do this with epistemological rules. For instance, the respective users of two inductive rules A and B that differ only in the value of a 'caution parameter' can agree that Rule A is more reliable but less powerful than Rule B; as a result, each realizes that a small shift in the relative value of reliability and power could lead to a preference for the other. In fact, the process of negotiating with people whose standards of evaluation differ from ours sometimes leads to a shift in our own standards (though of course such a shift is not evidence-driven). But though we sometimes negotiate or even shift standards in this way, we don't always: in dealing with a follower of the Reverend Moon, we may find that too little is shared for a neutral evaluation of anything to be possible, and we may have no interest in the evaluations that the Moonie gives. The fact that he gives them then provides no impetus whatever to revise our own evaluations, so the sceptical argument has no force from an evaluativist perspective.

Indeed, a main virtue of evaluativism is that it removes the force of most sceptical arguments. Most sceptical arguments depend on assuming that reasonableness is a factual property of beliefs or of rules, and on the understandable resistance to stripping away the normative nature of reasonableness by identifying it with a natural property like reliability (for rules; or being arrived at by reliable rules, for beliefs). Given the assumption and the understandable resistance to naturalistic reductionism, there is no alternative when faced with two radically different systems that positively evaluate themselves beyond (i) declaring them equally reasonable, (ii) postulating some mysterious non-natural property by which they differ, and (iii) saying that one is better simply by being mine (or more similar to mine). The second position seems crazy, and raises epistemological questions about how we could ever have reason to believe that a particular system has this property; the third position seems to strip away the normative force of reasonableness much as naturalistic reductionism did (indeed it could be regarded as a version of naturalistic reductionism, but one that uses chauvinistic natural properties); and this leaves only the sceptical alternative (i). Not a bad argument for scepticism, *if* one assumes that reasonableness is a factual property. Evaluativism provides the way out.

APPENDIX

Rules and Basic Rules

In the text I have tried to remain neutral as to whether a person's behaviour is governed by 'basic rules', but here I would like to argue that there is something to be said for supposing that this is so.

First a clarification: when I speak of someone 'following' a rule, what I mean is (i) that the person's behaviour by and large accords with the rule, and there is reason to expect that this would continue under a decent range of other circumstances; and (ii) that the person tends to positively assess behaviour that accords with the rule and to negatively assess behaviour that violates the rule. (In the case of epistemic rules, the 'behaviour' is of course the formation, retention, or revision of beliefs.) This is fairly vague, and the vagueness means that there is likely to be considerable indeterminacy involved in ascribing epistemic or other rules to a person: to ascribe a rule to a person is to idealize his actual behaviour, and idealizations needn't be unique. (I will discuss the significance of this shortly.) In any case, when I speak of rule-following I *don't* mean to suggest that the person has the rule 'written into his head'. There may be rules 'written into the head', but for those to be of use some part of the brain has to read them, and reading them is done by following rules; obviously these ones needn't be written in the head, on pain of regress.

In particular, when I imagined as a simple-minded illustration that we follow inductive rule (R) and that no evidence could lead us to change it, I certainly didn't mean to suggest that a person has something like my formulation of (R) 'written into his head', never to be altered by evidence. Even if some sort of formulation of the rule is explicitly written into the head, it might be very different from formulation (R). For instance, it might be that at a given time t what is written is not (R) but instead

> (R_t) If after t you have observed s_t ravens, and r_t of them have been black, you should believe to degree $(r_t+b_t)/(s_t+c_t)$ of any raven not yet observed that it is black,

where b_t and c_t are parameters representing the current bias, which changes over time. Following this sequence of rules is equivalent to following (R).[33] (If q_t is the number of ravens observed by time t and p_t is the number of them that have been black, then b_t and c_t are $j+p_t$ and $k+q_t$ respectively; since r_t and s_t are just $m-p_t$ and $n-q_t$ respectively, the equivalence is transparent.) 'Following this sequence of rules' might better be described as following the meta-rule (R*)

> Act in accordance with (R_t), *where the parameters b_t and c_t are obtained from earlier parameters by such and such updating process.*

But a psychological model could allow (R*) to be followed without being written into the head: the system is simply built to act in accordance with (R*), and to make assessments in accordance with it also. Again, no unchanging rule-formulation need be 'written into the head'.

A second clarification: not only don't I mean to suggest that the rule-formulations written into the head can't change over time, I don't mean to suggest that *the rules themselves* can't change as a result of observations: only that a person for whom that rule is fundamental can't recognize any observations as *evidence* undermining the rules. There are plenty of ways that the rules might change over time as a result of observations in a non-evidential way. Besides obviously non-rational changes (e.g. those produced by traumatic observations, or by computational errors), we might imagine changes made for computational convenience. Imagine a rule-formulation in the style of (R_t), where new evidence revises some parameters, but where the agent stores rounded off versions of the new values.[34] Over time, the values produced might start to vary considerably from what they would have been if the system had never rounded off. Here the rule-formulation changes to a *non-equivalent* rule formulation, on the impact of evidence; the rule itself changes. But this isn't a case where the accumulated observations serve as evidence against the old rule and for the new. (If we had started with a more

[33] As a model of what might be 'written into the head', the sequence of (R_t)s is far more plausible than (R): if (R) were what was written in it would require the agent to keep track of all the relevant evidence accumulated since birth, which is grossly implausible, in part because the computational requirements for storage and access would be immense. Still more plausible as a model is something 'in between' (R_t) than (R), where the agent doesn't need to remember all the evidence, but does remember some of it and retains a sense of what judgements he would make if something of the remembered evidence weren't in. (Indeed, something more like this is probably needed to handle assessments of our past inductive behaviour.)

[34] One might ask, why represent the meta-rule that the agent was following as the original (R*), rather than as a meta-rule that explicitly tells us to round off? I don't think this modified description of the agent would be incorrect, but neither do I think that a description of the agent as following the original (R*) would be incorrect: describing an agent as following a rule involves idealization of the agent's practices (especially when that rule is not explicitly represented in the agent, as it almost certainly wouldn't be for these meta-rules), and it's just a question of the extent to which one idealizes. Obviously, as the element of idealization of the agent's actual practices at revising beliefs lessens, the scope for arguing that the practices of revising beliefs can change lessens correspondingly.

complicated inductive rule than (R), there would have been more interesting ways for observations to lead to non-evidential changes in rules for purposes of computational simplicity.)

A less trivial example of how rules might change due to observations but not based on evidence arises when the rules are valued as a means of meeting certain goals (perhaps truth-oriented goals like achieving truth and avoiding falsehood). For there are various ways in which observations might cause a shift in goals (e.g. bad experiences might lead us to increase the weight given to avoiding falsehood over achieving truth), and thus lead to shifts in the rules for evaluating beliefs. But here too the shift in rules for believing isn't evidence-based, it is due to a change in goals. (It could also be argued that the basic rule in this example isn't the goal-dependent rule, but the rule about how beliefs depend on evidence *and* goals. This rule doesn't even change in the example described, let alone change as a result of evidence.)

Perhaps more important are cases where observations lead us to think up new methodological rules that had never occurred to us, and we are then led to use them on the basis of their intrinsic appeal. (Or maybe we observe others using them, and are led to use them out of a desire for conformity.) Here too it is transparent that the shift of rules is not due primarily to evidence against the old rules. Of course, on the basis of the new rules we might find that there is evidence against the old. But if the old rules didn't agree that it was evidence against them (and our resolution of the puzzle in Section 4 of the text says that they won't agree, if the rules are basic), then the decision to count the alleged evidence as evidence depends on an independent shift in the rule.

A third clarification: to assert that a person's inductive behaviour is governed by a basic rule is not to assert that there is a uniquely best candidate for what this basic rule is. To attribute a rule of inductive behaviour to someone is to give an idealized description of how the person forms and alters beliefs. For a variety of reasons, there need be no best idealized description. (The most important reason is that there are different levels of idealization: for instance, some idealizations take more account of memory limitations or computational limitations than do others. Also though I think less important, there can be multiple good idealized descriptions with the same level of idealization, especially when that level of idealization is high: since a description at a given highly idealized level only connects loosely with the actual facts, there is no reason to think it uniquely determined by the facts.) So there are multiple good candidates for the best idealization of our total inductive behaviour. Any such idealization counts any factors it doesn't take into account as non-rational. Insofar as the idealization is a good one, it is *appropriate* to take the factors it doesn't take into account as non-rational. The lack of a uniquely best candidate for one's basic rule is largely due to a lack of a uniquely best division between rational and non-rational factors.

With these clarifications in mind, let's turn to the issue of whether there are basic inductive rules. Since in attributing rules one is idealizing, really the only sensible issue is whether a good idealization will postulate a basic inductive rule (which might vary from one good idealization to the next). The alternative is an idealization that postulates multiple rules, each assessable using the others. But there is an obvious weakness in an idealization of the latter sort: it is completely uninformative about what the agent does when the rules conflict. There is in fact some process that the agent will use to deal with such conflicts. Because this conflict-breaking process is such an important part of how the agent operates, it is natural to consider it a rule that the agent is following. If so, it would seem to be a basic rule, with the 'multiple rules' really just default rules that operate only when they don't come into conflict with other default rules. Of course, this basic rule needn't be deterministic; and as stressed before, there need be no uniquely best candidate for what the higher rule that governs conflict-resolution is. But what seems to be the case is that idealizations that posit a basic rule are more informative than those that don't.

According to the discussion of the epistemological puzzle in Section 4, no rule can be empirically undermined by following that rule.[35] But if there are multiple candidates for one's basic inductive rule, it may well happen that each candidate C for one's basic inductive rule can be 'empirically undermined' by other candidates for one's basic inductive rule: that is, consistently adhering to a candidate other than C could lead (on certain observations) to a departure from the rule C. There's good reason to put 'empirically undermined' in quotes, though: 'undermining' C via C* only counts as genuine undermining to the extent that C* rather than C is taken as the basic inductive rule. To the extent that C is regarded as the basic inductive rule, it has not been empirically undermined.

I've said that the most important reason for the existence of multiple candidates for a person's basic inductive rule is that we can idealize the person's inductive practices at different levels. At the highest level, perhaps, we might give a simple Bayesian description, with real-number degrees of belief that are coherent (i.e. obey the laws of probability). At a lower level of idealization, we might give a more sophisticated Bayesian description, allowing for interval-valued degrees of belief and/or failures of coherence due to failures of logical omniscience. At a still lower level we might abandon anything recognizably Bayesian, in order to more accurately accommodate the agent's computational limitations. Eventually we might get to a really low level of idealization, in terms of an accurate map of

[35] That argument did not depend on an assumption that (candidates for) our basic inductive rules be deterministic. Suppose that our most basic rules dictate that in certain circumstances a 'mental coin-flip' is to be made, and that what policies one employs in the future is to depend upon its outcome. One can describe what is going on in such a case along the lines of (R) or (R*)—unchanging indeterministic rules, simply a new policy. In that case, obviously there is no change in the basic rules based on evidence, because there is no change in basic rules at all. Alternatively, one can describe what is going on along the lines of (R_t): the rules themselves have changed. But in this case, the indeterministic nature of the change would if anything *lessen* the grounds for calling the change evidence-based.

the agent's system of interconnected neurons, but using an idealization of neuron functioning. And of course there are a lot of levels of idealization in between. The rules of any one of these levels allow criticism of the rules of any other level as imperfectly rational: higher levels would be criticized for taking insufficient account of computational limitations, lower levels for having hardware that only imperfectly realizes the appropriate rules. But again, insofar as you somewhat arbitrarily pick one level as the 'level of rationality', then one's rules at 'the level of rationality' can't allow there to be empirical reasons for revising what is at that level one's basic inductive rule.[36]

References

Benacerraf, Paul (1973), 'Mathematical truth', *Journal of Philosophy* 70: 661–80.

Black, Max (1958), 'Self-supporting inductive arguments', *Journal of Philosophy* 55: 718–25.

Bonjour, Laurence (1998), *In Defense of Pure Reason* (Cambridge: Cambridge University Press).

Carnap, Rudolph (1963), 'An axiom system for inductive logic', in *The Philosophy of Rudolph Carnap*, ed. Paul Schilpp (La Salle: Open Court): 973–9.

Dummett, Michael (1978), 'The justification of deduction' in his *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press).

Field, Hartry (1994), 'Disquotational Truth and Factually Defective Discourse', *Philosophical Review* 103: 405–52.

—— (1996), 'The a prioricity of logic', *Proceedings of the Aristotelian Society* 96: 359–79.

—— (1998*a*), 'Mathematical objectivity and mathematical objects', in Stephen Laurence and Cynthia MacDonald (eds.), *Contemporary Readings in the Foundations of Metaphysics* (Oxford: Blackwell): 387–403.

—— (1998*b*), 'Epistemological Nonfactualism and the A Prioricity of Logic', *Philosophical Studies* 92: 1–24.

Friedman, Michael (1979), 'Truth and confirmation', *Journal of Philosophy* 76: 361–82.

Gibbard, Alan (1990), *Wise Choices, Apt Feelings* (Cambridge, Mass.: Harvard University Press).

Goldman, Alvin (1980), 'The internalist conception of justification', in *Midwest Studies in Philosophy*, v. 5, ed. Peter French, Theodore Uehling, and Howard Wettstein, (Minneapolis: University of Minnesota): 27–51.

—— (1988), 'Strong and weak justification', *Philosophical Perspectives* 2: 51–69.

Jeffrey, Richard (1983) 'Bayesianism with a human face', in *Testing Scientific Theories*, ed. John Earman (Minneapolis: University of Minnesota): 133–56.

[36] Presumably the rules at the very low levels in the hierarchy just described are in any reasonable sense beyond our control, whereas the higher-level rules should count as somehow 'in our control' (despite the fact that any changes made in the higher-level rules are due to the operation of the lower-level rules). One might want to stipulate that 'the level of rationality' is the lowest level of rules in our control. But 'in our control' is itself extremely vague, so this would do little to pin down a unique 'level of rationality'.

Kemeny, John (1963), 'Carnap's theory of probability and induction', in *The Philosophy of Rudolph Carnap*, ed. Paul Schilpp, (La Salle: Open Court): 711–38.

Kitcher, Philip (1983), *The Nature of Mathematical Knowledge* (Oxford: Oxford University Press).

Lewis, David (1971), 'Immodest inductive methods', *Philosophy of Science* 38: 54–63.

—— (1974), 'Lewis and Spielman on inductive immodesty', *Philosophy of Science* 41: 84–85.

Pollock, John (1987), 'Epistemic norms', *Synthese* 79: 61–95.

Putnam, Hilary (1963), '"Degree of confirmation" and inductive logic' in *The Philosophy of Rudolph Carnap*, ed. Paul Schlipp (La Salle: Open Court): 761–83.

Van Cleve, James (1984), 'Reliability, Justification, and the Problem of Induction' in *Midwest Studies in Philosophy*, v. 9, ed. Peter French, Theodore Uehling, and Howard Wettstein (Minneapolis: University of Minnesota): 555-67.