Philosophia (2008) 36:465–482
DOI 10.1007/s11406-008-9139-3

Relativism about Reasons

Nick Tosh

Received: 1 December 2007 / Revised: 17 April 2008 / Accepted: 28 April 2008 /
Published online: 6 June 2008

Abstract Historians must be sensitive to the alienness of the past. Insofar as they are concerned with their actors' reasoning, they must (through open-minded empirical investigation) *find out* how their actors thought, and not assume that they thought like us. This is familiar historiographical advice, but pushed too far it can be brought to conflict with rather weak assumptions about what historians must presuppose if they are to interpret their actors at all. The present paper sketches those assumptions, and argues that the influential 'Strong Program' in the Sociology of Scientific Knowledge (SSK) falls foul of them. We do not argue from the correctness of the assumptions to the falsity of SSK. Rather, we note the incompatibility, and then show how SSK theorists' tendency to *take interpretation for granted* blinds them—and perhaps their readers—to the existence of the conflict.

Keywords Relativism · Interpretation · Historiography · Sociology of Scientific Knowledge · Strong Program

Introductory Thought Experiments

One who believes that there are four plus three apples in the sack can be expected to believe that there are seven apples in the sack, because there being four plus three apples in the sack is a good reason to believe that there are seven. 'Expected' here is both predictive and normative ('England expects that every man will do his duty'). The man who believes that there are four plus three apples in the sack will (with high probability) believe that there are seven apples in the sack; and, whether or not he *in fact* does, he *ought* to believe that there are seven apples in the sack.

What sort of 'ought' is this? Is it merely an 'analyst's ought'—an expression of *our* normative standards, as when we say 'the Aztecs ought not to have practised

N. Tosh (✉)
Trinity College, Cambridge CB2 1TQ, UK
e-mail: njt26@cam.ac.uk

human sacrifice, the savages'? Or does it (also) purport to reflect the actor's norms? Are we saying that the actor ought by his own (or his own society's) lights to believe that there are seven apples in the sack? It is natural to feel that historians should concern themselves primarily with judgements of the latter sort; that the 'lights' in question can vary radically from culture to culture; and, therefore, that good historical scholarship requires a certain sort of open-mindedness. Though *we* condemn human sacrifice unreservedly, it would be the worst sort of anachronism to assume that the butchery at the Templo Mayor counted as deviant behaviour in Aztec terms. Of course, it is conceivable that those orchestrating the killings were violating their own society's standards; the point is simply that our sense of moral revulsion is poor grounds for thinking it likely. If there are good grounds to be had—and one's hunch in this case is that there are not—then we will obtain them only through careful empirical study of Aztec culture. Here, then, are two perfectly sensible historiographical intuitions. Historians should seek to understand the *actors*' norms; and they should train themselves to resist the (natural, human) tendency to underestimate the alienness of the past. The past must be allowed to speak for itself; it must be allowed to surprise us.

A historian who has taken these lessons to heart may feel inclined to question our opening remarks. 'The man who believes that there are four plus three apples in the sack ought to believe that there are seven apples in the sack.' This had better be an 'actor's ought', or the historian won't be interested; but if that *was* what we had in mind, the historian may want to register a protest along roughly the following lines. 'For *us*, playing *our* language game, there being four plus three apples in the sack is a good reason to think that there are seven apples in the sack; but whether the corresponding move counts as correct by the *actors*' lights is, ultimately, an empirical question. Let us go and *find out* how the actors thought, and not simply *assume* that they thought like us.' Notice that the predictive/explanatory potency of belief attributions is at stake here: if it is an open question whether the inference from P to Q counted as a good one by the actors' lights, it is an open question whether the actors' antecedent commitment to P is potentially explanatory of their subsequent commitment to Q. At first blush, this line of attack appears to be a straightforward application of the historiographical intuitions just surveyed. On closer inspection, however, it turns out to be problematic. Let us begin by spelling out the sort of scenario the historian has in mind.

We are invited to imagine a historical actor who believes that there are four plus three apples in the sack. Just what are we imagining here? Certainly not someone who utters, or is disposed to assent to, the sentence 'there are four plus three apples in the sack', for our actors may not speak English. But let us assume translation is possible. We are imagining someone disposed to assent to a sentence S that, when translated into modern English, yields 'there are four plus three apples in the sack'. Our translation manual further informs us that there is another sentence of the actors' language, T, which translates as 'there are seven apples in the sack'. The historian's 'empirical' question is whether, within the actors' language game, the inferential move from S to T counts as legitimate. If it turns out not to, we will have discovered that there being four plus three apples in the sack does not count, for the actors, as a good reason to think that there are seven. This, I take it, is the possibility our gedanken-historian wanted to leave open. But it is not entirely clear that it *is* a

possibility, even conceptually. Consider: we discover that, within the actors' language game, the inferential move from some sentence S to some other sentence T does not count as legitimate. That is obviously a conceptual possibility; and if we hold fast to certain assumptions about what S and T *mean*, we will get the result the historian wants. But there is no principled reason why we should hold fast. Translations are not God-given. Ultimately, we adopt the translation scheme that makes maximum sense of the actors' linguistic (and non-linguistic) behaviour. Translation and actors'-reasons-ascription are both empirical projects; they are both driven by the same sorts of data; and they cannot plausibly be pursued independently. These are vague observations, but the crucial point for our imagined case can be put rather crisply: that the inferential move from S to T counts as illegitimate by the actors' lights is strong—some would say decisive—evidence against the translation scheme assumed above.[1] In other words, the thought experiment is of doubtful coherence. Open-minded curiosity about the normative structure of the actors' language game—curiosity about which inferential moves, sentence to sentence, count as legitimate within that game—is all very well, but one can hardly plead *unlimited* open-mindedness on that front while making assumptions about translation. To assume that a particular translation scheme, a particular mapping from actors' utterances onto sentences of one's own language, is appropriate *just is* to assume that the actors' language game has a certain sort of structure: a structure compatible with the translation scheme in question.

The original claim was that the man who believes that there are four plus three apples in the sack ought, by his own lights, to believe that there are seven. It seemed highly plausible on its face. The historian's objection was that it smuggled in assumptions about the nature of the actor's 'lights' which might, in certain cases, prove unwarranted. To see why this objection is problematic, it is important to note that the original claim has the form—logically if not grammatically—of a conditional: *if* x believes that there are four plus three apples in the sack, *then* x ought (by his lights) to believe that there are seven. Commitment to that conditional for all x does not entail commitment to its antecedent for any particular x: in that sense one can be as open-minded as one likes. But if one *does* assert the antecedent for some x, one is *already* claiming something about the actor's 'lights': x must be playing a language game whose structure is such as to license the belief attribution in question. It is a fact about our language game—not an assumption about x's—that this belief attribution makes sense only given certain other interpretive commitments; and among those, I have suggested, is a commitment to the view that x ought, by his own lights, to believe that there are seven apples in the sack.

Why is this only a suggestion? Does the conditional hold, or doesn't it? Its prima facie plausibility makes it, I think, a useful crutch for our intuitions, but for the purposes of this paper I will not need to insist on its truth. And, indeed, some readers might on reflection want to question it. What of the tribe that counts to four but no further? Perhaps members of that tribe should not be regarded as having the concept

[1] To say that an inferential move is illegitimate by x's lights is not to say that it seems illegitimate to x; and to say that a move seems illegitimate to x is not to say that it is illegitimate by x's lights. (If, through sloppy calculation, I come to believe that 23+56 is 89, then I am wrong by *own* lights.)

plus, or at least as not having the concept *four plus three*. If that is right, then they pose no threat to our conditional; but the issue is one on which philosophers can be expected to differ. Our conditional isolates a particular belief—that there are four plus three apples in the sack—and declares it to have a particular consequence *by the believer's lights*, irrespective of his other commitments. Insofar as such claims are ever plausible, they are surely plausible in the case of arithmetic; but one who takes a Quinean view of analyticity might well reject them across the board. I do not want to take sides in that debate. Instead I will attempt to show, firstly, that rather weak assumptions—assumptions perfectly compatible with Quinean holism—are sufficient to constrain historians' open-mindedness vis-à-vis their actors' reasoning; secondly, that these assumptions can plausibly be seen as necessary presuppositions of interpretation; and thirdly, that they are nonetheless incompatible with a theoretical position—associated with Barry Barnes and David Bloor's 'Strong Program' in the Sociology of Scientific Knowledge (SSK)—that has become rather influential in history of science circles.

Empathic Reasoning: Interpretation and Explanation

Let us back off from the arithmetical example and consider the general case. We—the analysts, that is—regard P as a good reason for Q. To think that this commits us to the conditional *if the actors believed that P, then they ought by their lights to have believed that Q* is to collapse the distinction between our lights and theirs. That is obviously unacceptable. But we can hold on to the distinction without giving up the idea that an analyst's goodness-of-reason judgements play an essential role in his reconstruction of the actors' reasoning. Firstly, P's status as a good reason (by our lights) for Q may depend crucially on background assumptions. The actors may have believed R, though we do not; and we might *not* regard the *conjunction* P and R as providing good reason to believe Q. (We might even regard it as providing good reason not to believe Q, or to believe not-Q.) Call this the 'background problem'. Secondly, the chains of reasoning through which we think P can be seen to be a good reason for Q may all be very long and very complicated, so much so that the historical actors in question cannot possibly be expected to have been aware of them. On the principle that 'ought implies can', we may in such cases not want to say that the actors, who believed P, ought (by their lights) thereby to have believed Q. Call this the 'complexity problem'. The background problem and the complexity problem admit of a common solution: the analyst must empathise with his actors. As far as is possible, he must 'put himself in their shoes': he must imagine adopting all and only their beliefs, and devoting such time and resources to reflection upon those beliefs as he imagines were available to the actors themselves. When the analyst asks, within such a role-playing fantasy, 'what new beliefs ought I to acquire? Which of my current beliefs ought I to abandon? Which ought I to retain?', his answers induce, outside the fantasy, corresponding judgements upon the historical actors—judgements about what they ought, by their own lights, to believe (or cease to believe); and since, within the fantasy, the analyst is engaged in *reasoning*, it is clear that his goodness-of-reason judgements must play a crucial role.

We should not expect it to be easy to describe this role. Indeed, if by 'describe' we mean 'reduce to a well-defined set of rules and principles', there are good grounds for expecting it to be impossible. The grounds are these: empathic reasoning draws without limit on the resources of reasoning *simpliciter*; and the norms of rationality as such are *not* generally thought to be reducible to a well-defined set of rules and principles. For the purposes of the present paper that is a concession we are free to make and not a conclusion we are obliged to prove,[2] but it is worth noting in passing that we certainly cannot hope to understand *what the actors ought by their lights to believe* as given by anything much resembling the deductive closure of their actual belief set. The complexity problem is perhaps in itself a sufficient demonstration of this fact; here are two more. (1) To form the deductive closure of a set of beliefs is to extend it conservatively: it is to *add* entailments without taking anything away. But an empathic reasoner might perfectly well conclude that the actors ought to *abandon* their belief that P. (2) An empathic reasoner might also conclude that the actors, who in fact believe that P, ought to acquire the belief that not-P. But the conjunction of the actors' actual beliefs-at-a-time can entail the negation of one of them only if the conjoined beliefs are inconsistent—which, of course, they may very well be; but if they *are* inconsistent, their conjunction entails *anything*. The lesson is that the empathic reasoner cannot be replaced by a calculus of logical entailment.

We have said something tolerably precise about what empathic reasoning is not, but we have offered only a rather vague sketch of what it is. Furthermore, we have conceded that a precise picture is unlikely to be forthcoming. Where does that leave the claim that historians *cannot but* grant its legitimacy and relevance? The claim is plausible because it is plausible that empathic reasoning in the *same* loosely-specified sense is indispensable to interpreters: that is, that warranted interpretation is not possible without it. I devote the remainder of this section to fleshing out this view; along the way we will note some consequences for the theory of historical explanation.

That warranted interpretation is possible *at all* is not a claim that needs defending in the present context. That is because the possibility of warranted interpretation is a presupposition of any historical inquiry whatsoever: as practitioners of a 'human' and not a 'natural' science—of Geisteswissenschaft and not Naturwissenschaft—historians can hardly regard themselves as unentitled to make belief attributions. Of course, there are philosophically respectable ways of denying *anybody's* entitlement to interpretative claims; but blanket scepticism about interpretation is no more relevant to historians *qua historians* than blanket scepticism about induction is relevant to scientists qua scientists. When addressing historiographical questions, we take it for granted that historical actors can be interpreted. The question is what such interpretation involves.

The justification for the claim that empathic reasoning is indispensable to interpreters is the familiar thought that the latter must respect a *principle of charity*, seeking interpretative hypotheses that (ceteris paribus) minimise the gap between

[2]For a principled defence, see Child (1994), pp. 57–62, and works by Davidson and McDowell cited therein.

what the actors in fact believe and what they ought by their lights to believe. What the actors ought by their lights to believe depends, of course, on their environment and not simply on their actual beliefs. There is no suggestion here that empathic reasoning (as opposed to empathic 'responding') is a *sufficient* basis for interpretation. The suggestion is that it is *necessary*; that among the interpreter's indispensable rules-of-thumb is one that directs him to minimise, ceteris paribus, the 'coherence gap'—the gap between what the actors in fact believe and what they *thereby* ought (by their lights) to believe. This is the principle of charity in its *internal* aspect.

The ceteris paribus clause is difficult to spell out, but it is clear that it is required—and not simply because we are ignoring external (i.e., environmental) questions. Absent other constraints, an interpreter will always be able to make the coherence gap arbitrarily small in any number of mutually incompatible and individually unmeritorious ways. One could, for example, eliminate it altogether by taking the actors to have no beliefs at all. More interestingly, given any particular interpretative hypothesis, one can reduce the implied gap by resorting to ad hoc modifications of the hypothesis. Spelling out what it is, in general, for a modification to be ad hoc is no easier than spelling out the ceteris paribus clause (indeed, the problems are intimately related), but intuitively clear-cut illustrations are easily devised. The student who says '24 plus 17 equals 31' will, on the default interpretation, be wrong by his own lights; but should we so wish, we could articulate a new concept—following Kripke,[3] call it *quus*—which matches *plus* in all circumstances save, precisely, those of the student's assertion (when, of course, 21 quus 17 equals 31). It is a logical possibility that the student meant *quus* by 'plus',[4] but in most circumstances there will be little else to say in favour of the hypothesis. The lesson here is that the project of 'making sense' of the actors, even in its internal aspect, is not *simply* a matter of minimising the gap between what one takes the actors in fact to believe and what one thereby takes it that they ought to believe. But that is not to say that it isn't *in part* a matter of minimising that gap; it is merely to say that the interpreter will in general also have to juggle other, competing, desiderata. The interpreter who resorts too freely to quus-like hypotheses fails to make sense of his actors as *people*. (He fails sufficiently to respect a 'principle of humanity'.)

That coherence gaps need not be eliminated—that an interpreter might legitimately conclude that the actors at time t ought, by their own lights, to have believed P, though they in fact did not (or ought, by their own lights, not to have believed Q, though they in fact did)—has important consequences for diachronic explanation. Often, to explain an event is to show that (and how) it could have been *expected*. But to say that the

[3] Kripke (1982).

[4] At first sight it is tempting to deny that such wonky interpretations will be compatible with the actor's *other* behaviour. What if the student had said, in circumstances C, 'and for the benefit of eves-dropping philosophers, *obviously* I mean x + y to take the same value in all circumstances'? But in fact this is no help. Once we are willing to entertain wonky hypotheses about what somebody means, it doesn't matter what he *says*. To maintain the last wonky hypothesis, add another: that the student means *all circumstances* by 'all circumstances', *except* in circumstances C, when he means...

actors at time t ought, by their own lights, to have believed P *is* to say that the actors at time t could have been expected to believe P. Were this not the case, the idea that interpreters should seek, ceteris paribus, to minimise the gap between what they claim the actors did believe and what they thereby imply the actors ought (by their lights) to have believed *would make no sense*. The point of the principle of charity is not that unnecessarily uncharitable hypotheses are *unkind*, but that they are implausible: good interpretations minimise the violence done to our expectations. If the actors come to believe P at some later time t*, and it is still the case at and just before t* that they *ought*, by their lights, to believe P, then the latter fact provides, prima facie, a plausible explanation for the former: an expectation that happened not to have been fulfilled at t comes to be fulfilled at t*. (By the same token, interpretation as much as explanation may be diachronic: that the actors came to believe P at t* will sometimes be prima facie evidence that their antecedent beliefs were such as to make it the case that they ought, by their lights, to believe P.)

We are now in a better position to see why a certain sort of 'relativism about reasons' is—whatever its other merits—of dubious *historiographical* relevance. Here is the relativist line of thought I have in mind (I will present it as an objection to the above discussion). 'You have conceded already that the analyst can assess the logic of the actors' intellectual position only by imagining himself in their shoes. The analyst then, you go on to say, must ask "how ought *I* to revise my beliefs?" Asking and answering that question is all very well if the analyst wants to play the game of praising and blaming the actors—that is, praising those who thought like him and blaming those who did not. But now you suggest that his answers might legitimately ground *explanations* of the actors' behaviour, and here I must protest. It is no explanation of the *actors*' coming to believe P that the *analyst* feels obliged to do so when he imagines himself in their shoes!' We should admit at once that facts about the analyst do not explain facts about the actors. The question is not whether the analyst has a magical influence over the actors, but rather whether he is entitled to regard his own reasoning—in empathic mode—as to some extent *reliably indicative* of theirs. The relativist we are imagining insists on giving a negative answer. 'People reason in all sorts of ways. It is an empirical question whether the analyst's reasoning mirrors that of the actors; certainly he cannot *count* on it.' Now, of course, if 'mirrors' is read as 'is identical to', the relativist is perfectly correct: we have already conceded that an analyst might discover that the actors failed to believe what (on his interpretation) they ought by their own lights to have believed. Charity, as we saw above, has its limits. But that is not to say that it is dispensable; and if it is not dispensable for the purposes of interpretation, then one who is *committed to offering* interpretations can hardly pooh-pooh its deliverances when he turns his hand to explanatory projects. Expectations about the actors' beliefs formed on the basis of empathic reasoning are either to be regarded as reliable-up-to-a-point, or they are not. If, as seems plausible, the possibility of interpretation depends on their being regarded as reliable-up-to-a-point, and if, as seems incontestable, historians must offer interpretations of their actors, then presuppositions *necessary* for historical inquiry of *any* sort are *sufficient* to license the explanatory strategy rejected by the relativist. That is why the relativist's worry is of dubious historiographical relevance: there are strong grounds for thinking that historians cannot take it seriously while recognisably remaining historians.

It is worth noting what we have *not* assumed in this section. Firstly, our rough-and-ready account of empathic reasoning does not entail radical Quinean holism, but it is compatible with it. It therefore leaves open the question whether the proposition that x believes that there are four plus three apples in the sack strictly speaking *entails* the proposition that x ought (by his lights) to believe that there are seven. (One can, of course, give up the entailment claim while holding on to the thought that, on the whole, one who believes that there are four plus three apples in the sack can be expected to believe that there are seven.) Secondly, we have been concerned with *interpretation*—a procedure through which we form beliefs about the intentional states of others—and not with the metaphysics of intentional states as such. To say that the reliability (up to a point) of expectations formed on the basis of empathic reasoning must be presupposed by anybody who takes themselves to be in a position reliably to interpret historical actors is not, quite, to say that expectations formed on the basis of empathic reasoning *are* reliable. The legitimacy of Donald Davidson's 'transcendental' move from (something like) the former to (something like) the latter is in no way presupposed by the above discussion.[5] More generally, the *metaphysical* significance of questions about interpretability is a topic we have left untouched. There are fewer hostages to philosophical fortune in this section's account of interpretation than may at first meet the eye.

In any case, my ultimate aim in this paper is not to defend a particular philosophical theory of interpretation. I have sketched grounds for regarding as plausible the claim that the reliability of empathic reasoning is a necessary presupposition of interpretation; and SSK authors are indeed implicitly committed to rejecting that claim. However, our over-arching argument will *not* be that since the theory SSKers (implicitly) deny is true, SSK is false.[6] Rather, we will argue that SSK theorists' tendency to take interpretation for granted *blinds them to the existence of the issue*. To the extent that this blindness rubs off on their readers, and to the extent that the (rather orthodox) view of interpretation adumbrated above has any plausibility at all, the SSK position will seem more plausible than it ought to.

[5] 'Empathic reasoning' is not a phrase used by Davidson. However, his principle of charity 'directs the interpreter to translate or interpret so as to read some of his own standards of truth into the pattern of sentences held true by the speaker', and those standards include not only correctness (an external measure) but also *consistency* (Davidson 2001b, p. 148). Naturally speakers are not to be regarded as infallible: 'once the theory begins to take shape it makes sense to accept intelligible error and to make allowance for the relative likelihood of various kinds of mistake' (Davidson 2001a, p. 136). Responding to criticism from David Lewis, Davidson further concedes that one should allow for 'explicable error... right from the beginning [of one's interpretative project]' (Davidson 2001a, p. 282). Davidson's consistency principle, then, appears to be that an interpretative hypothesis is implausible insofar as the beliefs it ascribes are *unintelligibly* inconsistent—and this sits rather naturally with our talk of empathic reasoning. That Davidson attempts a *transcendental* move is easily shown: 'the only, and therefore unimpeachable, method available to the interpreter automatically puts the speaker's beliefs in accord with the standards of logic of the interpreter'; and 'what we, as interpreters, must take [the speaker's beliefs] to be is what they in fact are' (Davidson 2001b, pp. 150–151).

[6] I suspect that some such argument could be made to work, but proving so is beyond the scope of the paper.

SSK and 'Naturalism'

Sociologists of scientific knowledge have frequently been accused of treating scientific beliefs as determined, necessarily, by 'social' factors *rather than* by reasons.[7] It is a serious accusation, and if it could be made to stick it would be a lethal one. This much we can take for granted, for Barnes and Bloor do not dispute the accusation's seriousness: they dispute its justice; they purport not to recognise their critics' description of the SSK project. Barnes and Bloor insist only that genealogical explanations should be 'naturalistic' and 'causal': insofar as the actors' reasons *in fact helped cause* the actors' beliefs, the historian or sociologist of science is permitted, and might sometimes be obliged, to cite those reasons as causes. It is important to stress that this is no late concession. Barnes observed in 1974 that 'there is no necessary incompatibility between causes and reasons as explanations of action—indeed, reasons can be listed among the causes of action',[8] while in 1992 Bloor made a stronger claim: 'Far from being insensitive to the reasons that are given in the course of [a scientific] dispute, a sociological analysis *requires* them to be given a prominent role.'[9] Nonetheless, when Paul Boghossian wrote recently that 'it is impossible to see what would prevent our epistemic reasons from sometimes causing our beliefs',[10] he took himself to be objecting to SSK theorists in general and to Bloor's work in particular. On this point, it would seem, Barnes and Bloor can fairly claim to have been misrepresented by their critics. What accounts for this strange state of affairs? Why would so gifted a philosopher as Boghossian attribute to Bloor a position which Bloor has explicitly denied holding? The fault, I will argue, is not wholly Boghossian's. Bloor has an extremely thin notion of 'reason explanation'. It is, in fact, *so* thin a notion that philosophers of more conventional stripe could be forgiven for failing to recognise it.

We begin by quoting Bloor more fully:

> Far from being insensitive to the reasons that are given in the course of [a scientific] dispute, a sociological analysis *requires* them to be given a prominent role. The rival structures of reasons, and the differential credibility attached to those reasons, is the very thing that has to be explained.[11]

The rival structures of reasons is the very thing that has to be explained It looks as if the heralded 'prominent role' for the actors' reasons might simply be the role of *explanandum*—and if that is indeed the case, Boghossian and friends will be off the hook. Things aren't quite so simple, however. Bloor *is* committed to identifying 'structures of reasons', and he means *causal* structures. To identify x's belief that P as a motivating reason for x's belief that Q is to imply that the former belief helped

[7] See, e.g., Nola (1991) and Chapter 8 of Boghossian (2006).

[8] Barnes (1974), p. 75.

[9] Bloor (1992), p. 135. (Emphasis in the original.)

[10] Boghossian (2006), p. 117.

[11] Bloor (1992), p. 135.

cause the acquisition and/or the retention of the latter; and to imply that just is to commit oneself to a partial explanation. So reasons do explain. But Bloor's contention is that the historian or sociologist of science must always proceed one step further: he must explain why the structure of reasons in question was 'credible' from x's point of view. Now, one might think that this further stage of the explanation is a matter of identifying more of x's beliefs—beliefs which would explain x's taking P to be a reason for Q. Perhaps x believed $P \supset Q$. But that is *not* what Bloor means. In bringing in further motivating beliefs, we are simply describing a *larger* structure of reasons—and postponing the question of *why* that structure was credible to x. (One who insists on pressing on, hypothesising that x believed $(P\&(P \supset Q)) \supset Q$, is off on Lewis Carroll's famous regress.) Well, what of *empathic* reasoning? Will it do to imagine oneself adopting all and only x's beliefs at time t, 'see' that one ought to believe P, and thereby make sense of—*explain*—x's coming to believe P at t*? Bloor's answer to this question is well known: it certainly will *not* do. The historian or sociologist who proceeds in this way is destined to treat as 'self-explanatory'[12] such of his actors' reasoning as strikes him as rational, and to cast around for 'distorting' influences (social, psychological etc) to explain the rest. This is 'the asymmetrical, teleological picture'[13] to which SSK is firmly opposed. The analyst's explanatory project 'can, and should, be [pursued] without regard to the status of the belief as it is judged and evaluated by the sociologist's own standards'[14]—this is one version of the *symmetry thesis*—and what goes for individual beliefs goes also for structures of reasoning. The analyst must not allow his own sense of what is and is not rational to interfere with the empirical project of finding out what the actors in fact inferred from what. Bloor's position appears to be similar to that of the relativist we briefly imagined in the previous section. Our task now is to examine the intuitions Bloor takes to count in favour of that position.

Chief among them is the intuition that the historical and sociological study of science should be a 'naturalistic', even 'materialistic' enterprise. Explanations driven by what I have termed 'empathic reasoning' are not naturalistic in Bloor's sense. They amount, Bloor would say, to the following: x came to believe P at t* because, given his beliefs at t, that was the *rational* thing to do. But, he would continue, to describe the transition from one set of beliefs to another as 'rational' is to make a *normative* assessment of the transition. It is explanatory only insofar as one is committed to a 'teleological' view of history, according to which humans are providentially destined to achieve (or otherwise mysteriously guided towards) intellectual states that are in harmony with the demands of Reason. Bloor does not reject such a view out of hand—he describes it as an 'opposed metaphysical standpoint'[15]—but clearly expects that most of his readers will find it massively implausible. He sometimes presents it as a dualistic view, according to which

[12] See, e.g., Bloor (1991), p. 12.

[13] Ibid., p. 179.

[14] Barnes and Bloor (1982), p. 23.

[15] Bloor (1991), p. 12.

historians of thought must leave room for a guiding hand of Reason *over and above* the sorts of influences that can in principle be captured by physics:

> When we behave rationally or logically it is tempting to say that our actions are governed by the requirements of reasonableness or logic... It is as if we can transcend the directionless push and pull of physical causality and harness it, or subordinate it, to quite other principles and let these determine our thoughts.[16]

If such a picture is not to your taste, Bloor suggests, SSK is the natural alternative. SSK's 'symmetry requirement is meant to stop the intrusion of a non-naturalistic notion of reason into the causal story',[17] and among its defining commitments is the view that 'where previous analysts have had resort to a *non-naturalistic* and *normative* appeal to "reason" or "logic" or "meaning", it will prove fruitful to locate norms, conventions, traditions, and the like—that is, social processes.'[18] In addition to sociology, Bloor is in this context happy to admit the relevance-in-principle of empirical psychology, neuroscience, and indeed science of any sort. The key thing is the emphasis on *naturalistic* explanation.

Let us return to the method of empathic reasoning. To briefly recap: the analyst imagines himself adopting all and only x's beliefs at time t, concludes that he ought on that basis to believe P, and thereby (we have suggested but Bloor would deny) makes sense of x's coming to believe P at t*. What, exactly, is 'non-naturalistic' about this? Certainly there is an important sense in which the historian is treating his actors as people and not (simply) as things: one does not, after all, attempt to empathise with a continental plate or an electron. But if treating one's actors as people counts as breaking with naturalism, naturalistic historiography will be a contradiction in terms—and that can hardly be Bloor's point. We should focus instead on the suggestion that 'the intrusion of a non-naturalistic notion of reason into the causal story' is, somehow, the intrusion of a 'spooky' guiding force *over and above* the physical ones. If there were such a force then, since it is uncontroversial that human intentional states have physical consequences, physics would be inadequate even as a description of the *material* world; and while that is not quite a reductio ad absurdam—Bloor, indeed, is willing to grant that '[i]n its own terms the teleological model is no doubt perfectly consistent'[19]—it is an implication that even the least scientistic of historians might reasonably baulk at. Are there really grounds for thinking that 'previous analysts have had resort to a *non-naturalistic* and *normative* appeal to "reason"' *in anything like this sense*? Well, the accusation could no doubt be levelled at *some* previous analysts—unabashed Cartesian dualists; Divine Inspiration theorists; perhaps some Hegelians—but Bloor takes himself to be challenging contemporary philosophical orthodoxy. Among others, he is challenging Imre Lakatos, who took it for granted that one might sometimes explain the historical actors coming to believe P by showing that their earlier beliefs constituted

[16] Ibid., p. 8.

[17] Ibid., p. 177.

[18] Bloor (1992), p. 131. (Emphasis in the original.)

[19] Bloor (1991), p. 12.

good rational grounds for P. Bloor's claim is not merely that Lakatos exaggerates the power and scope of such explanations, but that from a naturalistic standpoint they are illegitimate in principle.[20] The view I will urge here is that Bloor finds this thought compelling chiefly because he fails clearly to distinguish between two levels of description: that of the *natural* sciences, and the intentional-state-attributing level characteristic of the *human* sciences.

It goes without saying that historians will never in practice have access to complete physical pictures of the events they study. Aristotle had some microphysical structure or other at every instant of his life, but we are unlikely ever to know the details. Suppose, however, that we *were* privy to a God's-eye-view of the physical facts, including all the physical facts about our historical actors; and suppose we saw that our actors' transitions, through time, from one physical state to another were entirely determined by physical law. That would certainly rule out guiding-hands-of-Reason of the *spooky* variety. We would have naturalistic explanations for every sound our actors uttered and for every mark they inscribed, and these explanations would be *complete* in the sense that they would cite antecedent states of affairs causally sufficient (given the laws) for the manifestation of the explananda. Bloor, perhaps, would think that a naturalistic tour de force of this sort would *leave no room* for Lakatos-style reason explanations: what more, he might ask, could one hope to explain? But this thought rests on a confusion. The far-fetched scenario we have imagined is *not* one in which the analyst has achieved a complete historical understanding of his actors. Indeed, the analyst has achieved no specifically *historical* understanding of his actors at all, because he has failed—or rather, has not attempted—to understand them *as people*. Given a diachronic physical description of the actors and their environment (whether partial or, *per impossibile*, complete), it will always be a non-trivial task to re-describe the actors in intentional terms; but in offering such a re-description, one is not necessarily attempting to fill explanatory gaps in the *physical* picture;[21] and it is in no sense a presupposition of the undertaking that the actors 'transcend the directionless push and pull of physical causality'.

But suppose there were no task of interpretation over and above the settling of the physical facts. Suppose, that is, that from our imagined God's-eye view the actors' physical states came 'pre-labelled' with intentional tags, and in particular came pre-labelled with tags specifying the actors' beliefs. Within such a fantasy, we can make some sense of Bloor's advice that we should resist 'the intrusion of a non-naturalistic notion of reason into the causal story': he is telling us, perfectly reasonably, that empathic reasoning, principles of charity and so on can and should be dispensed with. The actors had the beliefs they had. From our God's-eye perspective, we perceive those beliefs *directly*; and if on occasion certain bundles of beliefs *make no sense* to us, we may be surprised, but we will not be incredulous (we can hardly doubt the deliverances of a method that is, by hypothesis, infallible). Likewise for diachronic questions. Perhaps the actors' beliefs evolved through time in ways that

[20] Ibid., pp. 9–11.

[21] Although to the extent that the initial physical picture *is* incomplete *qua physical picture*, one may succeed in filling it in by making inferences that proceed *via* intentional state attributions.

to some extent make sense to us, and perhaps they did not; if the latter, we will have to swallow our pride and admit that the actors 'just thought differently'. We might, through careful empirical study, begin to understand some of the 'norms, conventions, traditions, and the like' governing (this to be understood naturalistically, in purely causal terms) the actors' acquisition and rejection of beliefs. Then we might be able to explain the actors' intellectual history at the *sociological* and not merely at the physical level, but there would still be no role for empathic reasoning. Once interpretation is taken for granted, naturalism of a certain sort and relativism of a certain sort can begin to look like natural bedfellows. But, of course, interpretation cannot be taken for granted, least of all by the historian. We cannot simply 'read off' the actors' beliefs, one by one, from the physical traces they have left us. We could not do so even granted a God's-eye view of the physical facts: granted such a view, our substantial remaining task would be to assign intentional tags to the various physical states of the actors in such a way that those sets of tags-at-a-time, and the transitions from set to set, rendered the actors comprehensible *as people*.[22] Bloor has nothing to say about interpretation; but he seems implicitly to assume that an analyst might in principle complete the 'empirical' project of chronicling the evolution of his actors' beliefs *before* raising the 'normative' question of whether he can make any sense of (1) the time-slices as (reasonably) coherent sets of beliefs, and (2) the transitions between them as (up to a point) rational by the actors' lights. It is important to note that this *is* an assumption rather than a conclusion for which Bloor argues; it is also significant that it is not explicitly articulated. If it is indeed the case that Bloor's position is incompatible with the not-very-controversial thought that interpretation is a matter, in part at least, of *making sense of the actors as reasoning creatures*, the most likely explanation is not that Bloor has an unorthodox theory of interpretation; it is that he has not thought specifically about *interpretation* at all.

Strings and Reasons

Bloor may not have much to say about interpretation, but he has plenty to say about meaning. With Barry Barnes, he has articulated a view that has become known as 'meaning finitism':

> Meaning finitism is the doctrine that meaning is always open-ended, and is created in a step-by-step fashion. Meaning follows usage, and so cannot exist ahead of it to guide it or act as an independent reality against which to measure usage.[23]

[22] In speaking of *assigning* intentional tags to physical states, I do not mean to imply that there is no objectively correct way of doing this. Let us suppose that there is. Then it will be an objective fact that a particular assignment of intentional tags to physical states constitutes the best interpretation of the actors; and there will be a straightforward sense in which the physical facts *taken together* determine the intentional ones. That is a metaphysical possibility we can afford to grant, thanks to the following epistemic point: to *know* the physical facts is not ipso facto to *know* which interpretation is objectively correct. Given the physical facts, the interpreter who wishes to *discover* the objectively correct interpretation can do no better than to follow the procedure described (sketchily) in the main text.

[23] Bloor (1996), p. 850.

We should immediately concede that the meanings of *words* depend on how they are conventionally used, and that conventional usage drifts with time along trajectories that cannot plausibly be seen as determined by prior meanings. None of that is controversial. But Bloor and other exponents of finitism want also to speak of *concepts*:

> Methods of accounting can connect any concept to any instance. Any particular concept-instance connection can be accounted genuine in any number of ways.[24]

> According to... finitism, we create meanings as we move from case to case. We *could* take our concepts and rules anywhere, in any direction, and count anything as a new member of an old class... We are not prevented by 'logic' or 'meanings' from doing this...[25]

We might now start to worry that the finitist intuition—'we *could* take our concepts anywhere'—rests upon a conflation of *quotation* and (more ambitious forms of) *interpretation*. Consider: if tomorrow we collectively decided to interchange the conventions governing our use of the words 'cat' and 'dog', we would thereafter (correctly) call cats 'dogs' and dogs 'cats'; but we would not thereafter believe (correctly or otherwise) that dogs purr and that cats bark. We would *say* 'dogs purr' and we would *say* 'cats bark'—we would hold true those *strings*—but to imagine our future selves using words differently is to imagine our future utterances standing in need, from our present point of view, of non-trivial *interpretation*. And of course similar points could be made about examples involving historical actors. This makes it difficult to make charitable sense of Barnes and Bloor's more extreme finitist pronouncements. More importantly, it is evidence in favour of our hypothesis that these authors are uninterested in, and perhaps insensitive to, questions of interpretation.

There is other evidence. Most philosophers who write on the general topic of meaning make some use of foreign language examples, and most explicitly address the issue of translation. Barnes and Bloor do neither. Furthermore, they frequently flip-flop between talk of concepts and talk of words. For example, Barnes summarises one of the intuitions driving finitism as follows: 'People... are not under the control of their discourse or their own *verbal artefacts*: the relationship is the other way round.'[26] Barnes, Bloor and Henry, discussing the problem of explaining the actors' acts of classification—a problem which is intended to serve as a paradigm for explaining concept application in general—stress that it is a 'basic tenet of the finitist account' that the causes cited by the analyst 'may not include the alleged intrinsic power of words and verbal formulations.'[27] We could indeed take our *verbal formulations*—strings of written or spoken symbols—anywhere. Bloor is quite right that neither 'logic' nor 'meanings' prevent us. There *is* no logic of strings *qua strings*; and the

[24] Barnes (1981), p. 482.

[25] Bloor (1997), p. 19. (Emphasis in original.)

[26] Barnes (1981), p. 481. (My emphasis.)

[27] Barnes et al. (1996), p. 119. For slither between 'proposition' and 'verbal formulation', see Bloor (1997), p. 59.

'meanings' they carry within a particular community at a particular time depend on re-negotiable conventions.

There are, then, grounds for suspecting that Bloor might be prone to conflate claims about what the actors believed with claims about what strings they held true. The suspicion that he fails clearly to distinguish between natural scientific and intentional-state-attributing levels of description, while not equivalent (to say that the actors held true a string is already to attribute an intentional state), is at least related: in both cases our worry is that Bloor has failed to appreciate what is involved in interpretation. With this in mind, we should return to the topic of reason explanation. That the actors *held true the strings* S_1, S_2,... and S_n is, in and of itself and in the absence of further information, *no* reason to expect the actors to have held true some further string T; nor could it ever be counted a satisfying explanation of their coming at some later time to hold true T. This is true whatever meaning (if any) the strings in question have in the *analyst's* language; in particular, it is true even if it so happens that the analyst would hold true T if he held true S_1, S_2,... and S_n. But if we imagine adding some sociological meat to the formal skeleton, the picture changes. That it is the case *both* that the actors held true the strings S_1, S_2,... and S_n *and* that certain 'norms, conventions, traditions' obtained in their society—norms, conventions and traditions related to the use of strings—might very well be a reason to expect the actors to have held true some further string T, and might very well be a satisfying explanation of their coming at some later time to hold true T. What we have just seen is emphatically *not* that reason explanation is impotent until 'fleshed out' by the sociologist, because reason explanation does not proceed at the level of strings: it proceeds at the level of beliefs. But one who is careless of the quotation/belief attribution distinction might *think* we have just seen that reason explanation is impotent until fleshed out by the sociologist.

As evidence that Bloor thinks something rather like this, I will quote a passage in which he is talking about *rule-following*. One preliminary comment is in order. Rule-following explanations are closely related to reason explanations: both purport to explain what the actors in fact did in terms of what they should have done; and in both cases, the quotation/attribution distinction is crucial. That the actors were following (or attempting to follow) a rule they happened to label 'add 2' is not, in and of itself and in the absence of further information, any explanation of their taking 8 to be the next number in the series 0, 2, 4, 6; but that they were following (or attempting to follow) the rule *add 2* is. Bloor writes:

> Philosophers... have typically tried to identify some feature of [the] previous cases that makes it necessary that the next step in following the rule, or the next application of the concept, should be thus and so. Wittgenstein says they have failed to locate the source of this necessity because they have been looking in the wrong places. They have appealed to abstract 'meanings', or 'definitions', or 'interpretations' or 'conceptual content' or 'what the rule requires'. All such 'formal' specifications (as they might be called) will fail because they represent a justification that sets in train an infinite regress of further, formal specifications. Thus if a rule depends on an interpretation then the interpretation demands an interpretation. This, said Wittgenstein, could lead anywhere. Anything could be deemed an interpretation of anything else. He concluded that

> the real determinants of the next application, and the real sources of the discrimination between correct and incorrect steps and applications, were not to be found in the realm of formal specifications and justifications but amongst the totality of contingencies that impinge on the episode. He was not saying that the move to the next case was undetermined. Rather, the determinants lie around or behind the formal specifications but do not appear in or amongst them.[28]

Well, do the determinants *not* appear among the formal specifications? It depends what we mean by 'formal specifications'. If we mean the analyst's specifications of the *strings* which he thinks the actors held true or tried to 'obey', there *is* an important sense in which the determinants lie chiefly 'around or behind the formal specifications'. If we say that the actors were attempting to follow a rule which they labelled 'add 2', but make no further assumptions, we are in no position to explain anything; and if we try to rectify matters by additionally citing the strings the actors took to spell out what they meant by 'add 2', we are indeed starting ourselves off on an infinite regress. But that is not a very interesting observation. If, by contrast, the analyst's judgement that the actors were following the rule *add 2* (NB the absence of quotation marks) counts as a formal specification, the regress worry does not arise in a form that ought to trouble historians; for historians are entitled to assume that they have a grasp of their own language, and that is all empathic reasoning requires ("if I believed P, Q and R and was trying to follow the rule *add 2*, then I ought to write down '8, 10, 12...'"). My point here is not that the Kripkensteinian meaning sceptic could not dispute the claim that historians have a grasp of their own language; my point is that sceptical doubts of this sort are not historiographically relevant. They are too general. Since they threaten every historical claim *equally*, they cannot provide grounds for favouring any particular historical claim over any other.

Let us return to Bloor's text:

> This result is important for sociologists of knowledge because, amongst the actual determinants and contingencies at work, there will, and must always be, social processes. These were explicitly identified by Wittgenstein himself as customs, conventions and institutions. Clearly, Wittgenstein's argument insinuates the social into the most intimate and pervasive of all cognitive processes.[29]

Our question is: does this 'insinuation of the social'—Bloor elsewhere has written that finitism 'shows the social character of that most basic of all cognitive processes: the move from one instance of concept application to the next'[30]—mean that reason explanations are impotent until fleshed out by the sociologist? Our answer turns on the string-held-true/belief distinction. Recall that talk of strings-held-true was not in itself sufficient to fashion even string-level *ersatz* reason explanations; we needed to make reference to Bloor's 'norms, conventions, traditions' *in addition*. But what

[28] Bloor (2007), pp. 211–212.

[29] Ibid., p. 212.

[30] Bloor (1991), p. 165.

about talk of the actors' *beliefs*? One way to approach the question is to note that reason explanations (in terms of beliefs) can be regarded as *inducing* string-level explanations. If one interprets the actors as having believed that P on the evidence of their holding-true of string S, and as having believed that Q on the evidence of their holding-true of string T, then *in taking* P to have been a motivating reason-by-their-lights for Q one is taking the 'norms, conventions, traditions' governing their string use to be of a certain sort: such as to license their move from the holding-true of S to the holding-true of T. Once again, sceptical worries about our entitlement to such commitments can be evaded by pointing to presuppositions necessary for any historical inquiry; for simply by taking the actors to have believed that P, one is taking them to have 'played' a string-manipulating game[31] of a certain sort: one whose structure is such as to warrant the attribution to them of the belief that P. The thought, then, is that it is misleading to think that reason explanations are, in principle, impotent until fleshed out by sociologists. Explanations are not *reason* explanations at all, even skeletal ones, unless they traffic in belief attributions; and, to the extent that assumptions about 'norms, conventions, traditions' are *sociological* assumptions, any explanation that involves propositionally complex belief attributions (of the sort that would not be plausible with respect to non-linguistic creatures) is already sociologically 'meaty'.

Nobody will deny that questions about what the actors ought by their lights to have believed are empirical questions. For the sake of argument we can further concede that answering such questions is always in part a sociological project. But this is *not* to concede that determining what the actors ought by their lights to have believed *given what they did in fact believe* is always a sociological (or even an empirical) project. That much stronger thesis, and the 'relativism about reasons' that is its natural companion, will seem compelling chiefly to those who are willing to conflate belief attribution with mere quotation.

Acknowledgements I am grateful to Daniel Elstein, Adrian Haddock, Nick Jardine, Martin Kusch, Florian Steinberger and seminar audiences in Cambridge and Belfast for valuable feedback on talks and drafts ancestral to this paper. I am grateful also for the constructive criticism of an anonymous referee.

References

Barnes, B. (1974). *Scientific knowledge and sociological theory.* London: Routledge.
Barnes, B. (1981). On the 'hows' and 'whys' of cultural change (response to Woolgar). *Social Studies of Science*, *11*, 431–498.
Barnes, B., & Bloor, D. (1982). Relativism, rationalism and the sociology of knowledge. In M. Hollis & S. Lukes (Eds.), *Rationality and relativism* (pp. 21–47). Oxford: Blackwell.
Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific knowledge*: *A sociological analysis*. Chicago: Chicago University Press.
Bloor, D. (1991). *Knowledge and social imagery.* Chicago: Chicago University Press.
Bloor, D. (1992). Ordinary human inference as material for the sociology of knowledge. *Social Studies of Science*, *22*, 129–139.
Bloor, D. (1996). Idealism and the sociology of knowledge. *Social Studies of Science*, *26*, 839–856.

[31] He we use 'string' in a broad sense, to cover strings of sounds as well as strings of (written) symbols.

Bloor, D. (1997). *Wittgenstein: Rules and institutions*. London: Routledge.
Bloor, D. (2007). Ideals and monisms: Recent criticisms of the Strong Programme in the sociology of knowledge. *Studies in History and Philosophy of Science*, *38*, 210–234.
Boghossian, P. (2006). *Fear of knowledge*. Oxford: Oxford University Press.
Child, W. (1994). *Causality, interpretation, and the mind*. Oxford: Oxford University Press.
Davidson, D. (2001a). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
Davidson, D. (2001b). *Subjective, intersubjective, objective*. Oxford: Oxford University Press.
Kripke, S. (1982). *Wittgenstein on rules and private language: An elementary exposition*. Cambridge: Harvard University Press.
Nola, R. (1991). Ordinary human inference as refutation of the Strong Programme. *Social Studies of Science*, *21*, 107–129.