An Introduction to Contemporary Epistemology

JONATHAN DANCY

Basil Blackwell

2
Knowledge

2.1 THE TRADITIONAL ACCOUNT

The standard account of knowledge, around which all recent work has been done, defines knowledge as justified true belief; it holds that a knows that p if and only if

1 p,
2 a believes that p,
3 a's belief that p is justified.

Because there are three parts to this definition it is called the tripartite definition or the tripartite account; it defines propositional knowledge, knowledge that p; it does not define knowledge by acquaintance as in 'a knows James' nor knowledge-how, e.g. knowledge how to ride a bicycle, unless these can be shown to reduce to knowledge-that.

The tripartite definition has obvious attractions. The first clause, that if a knows that p then p is true (which can be read as $Kap \rightarrow p$), is normally seen as stipulative. The second clause, that if a knows that p then a believes that p (we can read this as $Kap \rightarrow Bap$), seems minimal, and the third, that if a knows that p then his belief that p is justified ($Kap \rightarrow JBap$), is there in order to prevent any lucky guess from counting as knowledge if the guesser is sufficiently confident to believe his own guess. It is worth noticing, however, a consequence of this justification of clause 3; this is that a belief is not generally considered to be justified by the mere fact that it is true, for otherwise clause 3 would be unnecessary. If I decide on the toss of a coin which investment will provide the greatest yield, and fortunately turn out to be right, we suppose that my choice is vindicated by the outcome perhaps, but not justified by it; I had no real justification for making the choice I did. (Alternatively we

could distinguish between two forms of justification, justification before the event and justification after it, and run the tripartite definition in terms of the former; but then the question would be whether these really are two forms of the same thing.)

What are the problems for the tripartite definition? One might think that clause 2 is insufficient: to believe that p is not so strong as to be certain that p, and to know one must be certain, not just believe.

The best reason for wanting some account of certainty in our analysis of knowledge is that people are rightly hesitant to *claim* knowledge when they are less than certain. This hesitation seems to be due to something about what knowledge is, and there is no obvious way to explain it if knowledge is as the tripartite conception claims it to be. Thus although it is commonly suggested that the notion of certainty is relevant to the analysis of *claims* to knowledge, but not to the analysis of knowledge itself (e.g., in Woozley, 1953), this leaves us with no method of explaining why certainty should be required before one can claim knowledge when it is not required for knowledge itself, i.e., for the existence of what one is claiming.

Since we are going to discover other reasons for rejecting the tripartite definition, we have no reason to pursue this point here. The moral to be drawn is that if we are to give an account of knowledge which does not include a requirement of certainty, our account should make room for the notion of certainty somewhere; if it sees certainty as a requirement for a knowledge claim it needs to be able to explain in its own terms why that should be so.

But why should we be reluctant simply to change clause 2 to 'a is certain that p'? The answer is that we are prepared, in circumstances that are not particularly unusual, to allow that someone does in fact have knowledge when that person is so far from certain that he would not claim the knowledge himself. The classic example offered is that of the diffident schoolboy, who has learnt the dates of, say, the English kings the previous night but who is so alarmed by his hectoring schoolteacher that he becomes completely unsure that the answers that suggest themselves to him under questioning are in fact the right ones. Supposing, however, that those answers are correct, would we not allow that he knows them, even though he himself might not make that claim? And surely our reasons for allowing this are close to those suggested by the tripartite definition; he has the right answer, and not by luck.

There is a weakness in this appeal to the diffident schoolboy,

which concerns clause 2 again. To the extent that the schoolboy is less than certain of the answers that occur to him, can we allow that he still believes them? If we are not careful, use of this example to defuse the pretensions of a certainty condition will result in our losing the belief condition we were trying to defend.

2.2 GETTIER COUNTER-EXAMPLES

Henry is watching the television on a June afternoon. It is Wimbledon men's finals day, and the television shows McEnroe beating Connors; the score is two sets to none and match point to McEnroe in the third. McEnroe wins the point. Henry believes justifiably that

1 I have just seen McEnroe win this year's Wimbledon final.

and reasonably infers that

2 McEnroe is this year's Wimbledon champion.

Actually, however, the cameras at Wimbledon have ceased to function, and the television is showing a recording of last year's match. But while it does so McEnroe is in the process of repeating last year's slaughter. So Henry's belief 2 is true, and surely he is justified in believing 2. But we would hardly allow that Henry knows 2.

This sort of counter-example to the tripartite account of knowledge is known as a Gettier counter-example, after E. L. Gettier (1963). (I owe this particular example to Brian Garrett.) Gettier argued that they show the tripartite account to be insufficient; it is possible for someone not to know even when all the three clauses are satisfied.

Gettier here is not quarrelling with any of the three clauses. He allows that they are individually necessary, and argues only that they need supplementing.

It is worth formalising the situation, for reasons which will emerge later. Reading 1 as p and 2 as q, we have:

$\sim p$, Bap, JBap, $p \rightarrow q$, JB$a(p \rightarrow q)$, q, Baq, JBaq.

So a Gettier counter-example is one in which a has a justified but

false belief by inference from which he justifiably believes something which happens to be true, and so arrives at a justified true belief which is not knowledge.

What response should be made to these infamous but slightly irritating counter-examples? There seem to be three possible routes:

1   find some means to show that the counter-examples do not work;
2   accept the counter-examples and search for a supplement to the tripartite analysis which excludes them;
3   accept the counter-examples and alter the tripartite analysis to suit rather than adding anything to it.

The remainder of this section is concerned with the first route.

On what principles of inference do these counter-examples rely? Gettier himself exposes two. For the examples to work, it must be possible for a false belief still to be justified; and a justified belief must justify any belief which it implies (or is justifiably believed to imply). This last is just the principle of closure PC^j mentioned above in the discussion of scepticism (1.2). So if we could show PC^j false this would have the double effect of undermining the Gettier counter-examples and (part at least of) the first sceptical argument. It might be possible, however, to construct new variants on the Gettier theme which do not rely on inference or on an inference of this sort, as we shall see in the next section. and if so no complaints about PC^j or other principles will be very effective.

One thing we cannot do is to reject Gettier counter-examples as contrived and artificial. They are perfectly effective in their own terms. But we might reasonably wonder what point there is in racking our brains to find an acceptable definition of 'a knows that p'. Is this more than a mere technical exercise? What, if anything, should disconcert us if we cannot come up with a trouble-free definition? Many of the innumerable papers written in response to Gettier give the impression that responding to Gettier is a kind of private philosophical game, which is of no interest except to the players. And hasn't Wittgenstein shown us anyway that a concept can be perfectly healthy without being definable, arguing that there need be no element common to all instances of a property (e.g. instances of knowledge) other than that they are instances (e.g. that they are knowledge)? (Cf. Wittgenstein, 1969b, pp. 17 – 18, and 1953, §§ 66 – 7.) So what on earth could depend on our success or failure to discover necessary and sufficient conditions for knowledge?

In many ways I sympathize with the general tenor of this

complaint, as may quickly become apparent. What sustains me in the search for a response to Gettier is the feeling that it may be possible to find an account of what knowledge is which will have a substantial effect on what we are to say about justification in later parts of this book. This could happen in either or both of two ways. We might find an account of what knowledge is which would suffice to undermine crucial sceptical moves, and hence confirm the possibility that some of our beliefs are justified; the account to which I give tentative support in chapter 3 has pretensions in this direction. Or we might hope to define justification in terms of knowledge. For instance, we might suppose that a belief is justified iff in certain circumstances (to be spelled out) it would be knowledge. (Jennifer Hornsby gave me this idea.) In the meantime we must consider some accounts of knowledge which seem less fruitful.

2.3 RESPONSES TO GETTIER

I rather obviously avoided, so far as I could, offering even the most tentative diagnosis of the defect in the tripartite analysis which Gettier exposed. This is because the different responses to Gettier all stem from different diagnoses of the way in which the tripartite analysis is lacking; once we know what is missing, it should be quite a simple matter to add it.

The Presence of Relevant Falsehood

The most obvious diagnosis is simply that the initial belief that p, from which the true justified belief that q is inferred, is false. So we might add to the tripartite analysis the fourth condition that nothing can be known which is inferred from a false belief, or from a group of beliefs of which one is false.

This simple suggestion has two defects. First, variants on the Gettier theme can be written in which, though there is falsehood, there is no inference. Suppose that I believe that there is a sheep in the next field because of what I see. I am not inferring from what I see that there is a sheep in the field; I take myself simply to see that there is one. The animal I see is a large furry dog, but my belief is not false, because there is a sheep there too, unknown to me, hidden by the hedge. Here we might admit that my belief is true and justified but refuse to grant that I know there to be a sheep in the field. (This example comes from Chisholm, 1977, p. 105.)

A reply might be that surely I am inferring that I see a sheep in

the field from my knowledge of my own present sensory states. This reply raises large issues; but chapter 5 contains a lengthy argument that if there is any non-inferential knowledge, some of it concerns things other than our sensory states – so why not sheep, for instance?

The second defect is that the suggestion is too strong and is likely to make it impossible for any of us to know anything at all. As we shall see, this is a danger with a number of responses to Gettier. In the present case, we all of us suffer from numerous false beliefs which have some role in our inferential processes, and so on this suggestion none of our present true justified beliefs would count as knowledge.

To eliminate these defects we must remove the reference to inference and tighten up the relation specified between the false beliefs and the true justified ones which are not to count as knowledge. Thus we could simply require an absence of relevant falsehood. This would get round the example of the sheep in the field because I presumably believe (falsely) that the animal I can see is a sheep even though this belief is not used in inference. But as a suggestion it seems rather to name the difficulty than to solve it: which false beliefs are to be counted as relevant?

An answer might be that a false belief that p is relevant in the required sense if, had the believer believed instead that $\sim p$, his belief that q would cease to have been justified. Not all false beliefs are relevant in this sense. Some will be so distant or insignificant that whether one believes them or their opposite would have no effect on what one believes here. For instance, among the beliefs in virtue of which I claim to know that Napoleon was a great soldier there may be one which is false, but which is so insignificant that my justification for believing that Napoleon was a great soldier would survive my changing my mind on that particular point. Such a false belief would not be relevant in our present sense.

But the new account faces difficulties, which can best be illustrated by an example. Suppose that I expect a colleague to give me a lift home this evening, but that her car has a flat battery; this won't stop us, however, because a friend's car is parked conveniently near with some jump leads which we can use to get her car to start. I now believe that she will give me a lift this evening, and this belief is justified. Do I know that she will give me a lift? The requirement that there be no relevant false beliefs suggests plausibly that whether I know depends on what other beliefs I have. But this suggestion raises difficulties. If, for instance, I merely believe

1 she will give me a lift home this evening,

I may be allowed to know this, but if I believe both 1 and

2 her car battery is not flat,

I may not, since I have a relevant false belief. But if, as well as believing 1 and 2, I happen also to believe

3 there is a friend's car conveniently near with jump leads,

then this apparently gratuitous belief makes it the case again that I know that I will be given a lift. For if I had believed the opposite of 2, my belief 1 would not be justified unless I also believe something like 3. It seems then that our present suggestion has the effect that whether I have knowledge will depend commonly on which other apparently gratuitous beliefs I may have. There is something unsatisfactory about this, and more work needs to be done to defend the account against complaints of this sort.

Defeasibility

A slightly different approach diagnoses the Gettier counter-examples as arising because there are some truths which would have destroyed the believer's justification had he believed them (cf. Lehrer and Paxson, 1969; Swain, 1974). Thus, for instance, suppose that Henry had believed that he was watching a recording of last year's final, as he was; in that case, his justification for his belief that p and thus by PC^j for his belief that q would have been destroyed. The suggestion then is that a fourth clause be added requiring that there be no other truth such that Henry's believing it would have destroyed his justification for believing that q. This is the defeasibility suggestion; we require for knowledge that the justification be indefeasible, i.e. that the addition of further truths should not defeat it.

This will not imply that a false belief will never be justified, since the suggestion is that although some beliefs are defeasibly justified, we require indefeasible justification for knowledge. However, it is in danger of rendering the first condition for knowledge ($Kap \rightarrow p$) redundant. It looks as if a false belief could never be indefeasibly justified since there would always be some truth (even if only the negation of the false belief) whose addition would destroy the justification. But perhaps this is a strength in the theory rather than a weakness, since the new quadripartite analysis will have a

coherence that was previously lacking; it provides an explanatio
in the fourth clause of what was before included by mere stipulation,
that knowledge requires truth.

The defeasibility suggestion could be said to provide an extension
of the earlier requirement that there be no relevant falsehoods; we
now look beyond those propositions actually believed by the believer
to propositions which would have an effect if they were believed.
But this extension is no real advantage. The sort of difficulty facing
the notion of defeasibility can again best be illustrated by an
example. Thus, perhaps, I believe that my children are at present
playing in the garden at home, and I have good reasons for this
belief. However, unknown to me, a neighbour rang up after I left
home this morning to invite the children round for the morning.
And if I had known this my justification for believing that they
are playing at home would be defeated, because I also believe that
they normally accept such invitations. However, my wife has become
concerned about the health of one of them and refused the invita-
tion. Do I know that my children are playing in the garden at home?
If your intuition is that I do, you must reject the defeasibility
criterion as formulated at present. If it is that I do not, on the
grounds that had I heard about the invitation my justification would
have been defeated, you have a duty to give some account of why
the (unknown to me) truth that my wife has refused the invitation
does not somehow redress the balance. Either way the defeasibility
proposal needs to be altered.

The problem seems to lie, as it lay for the requirement that there
be no relevant falsehood, in the way in which new true beliefs can
be added piecemeal and overturn the existing justification, while
there remain yet further truths waiting in the background to over-
turn the overturning. First we want to ask anyway whether there
isn't likely always to be some truth which, if it alone were added
and all others excluded, would defeat my justification. Even if this
won't always happen, it will certainly happen often enough for the
range of my knowledge to be considerably reduced, and this itself
is some sort of an objection. Second, we need to find a way to
counter the way in which the piecemeal addition of further truths
seems to switch me into knowledge and out again.

We might achieve the second task by altering our account of
defeasibility so that instead of talking about some one other truth
(which caused the problem of piecemeal addition) we talk about
all truths whatever. Thus we could require as our fourth condition
that our justification would remain even when every truth is added to

our belief set, all at once. This new notion of defeasibility seems to allow (probably) that I now know that my children are playing in the garden, because the second added truth negates the defeating powers of the first. But there remain problems for this new notion of defeasibility. First, in talking of adding all truths at once we seem to have moved firmly into the realm of fiction. Indeed, do we have any suitable conception of 'all the truths'? Second, it seems on this criterion that we shall never have more than the slenderest of reasons to believe that we know something; for in believing this we are believing that when all the truths are in, our justification will remain, and it looks as if much more is required to support that belief than is required to support an ordinary claim to know.

Reliability

A different approach diverts our attention away from the relation between the proposition claimed as knowledge and other false beliefs which should have been true or other truths which should have been believed. It is sometimes suggested that a justified true belief can be knowledge when it derives from a reliable method (see Goldman, 1976; Armstrong, 1973, ch. 13; Swain, 1981). In the Gettier example, Henry does know that the Wimbledon final is being played that afternoon; this justified true belief derives from the reliable method of reading the newspapers, which are normally right about this sort of thing. However, his belief that q clearly derives from a method that is less than reliable. It would have led him badly astray here, had McEnroe suffered an unexpected lapse and succumbed to the efforts of Connors for once.

The reliability approach can be made more elaborate; in some ways it is closely related to the causal approach considered next, because we are clearly owed an account of what reliability is, and a causal answer is tempting (see, e.g., Goldman, 1979). However, we can already see difficulties for any variation on this approach. It is in danger either of making knowledge impossible or of walking straight into one of our sceptical arguments.

We may mean by 'reliable' that a suitable method, if properly followed, is perfectly reliable and never leads to a false belief. But, quite apart from the general difficulty of distinguishing between a defect in the method and a defect in the manner in which the method has been applied, it seems improbable that there are any perfectly reliable methods of acquiring beliefs. Man is fallible, and his fallibility is shown not just in the manner in which the methods are used but in the belief-gathering methods available to him. Hence

if knowledge requires an infallible or perfectly reliable method, it is impossible.

But if we retreat from the notion of perfect reliability and require only that the method be generally reliable, we invite sceptical arguments of our second type. How is it that a method which has failed elsewhere in relevantly similar circumstances suffices to yield knowledge this time? If we had any hope that our eventual account of knowledge would help us to reject the sceptical arguments, this particular account seems to make matters worse rather than better. Of course this won't show that the account is wrong. It may be that the correct account of knowledge does unfortunately give the sceptic the opening he is looking for. But we should not accept that this is how things are until we are convinced that there is no other account of knowledge which offers the sceptic less leverage. We can still hope for one which makes life harder rather than easier for him.

A final retreat would be to require only that the method be reliable *this* time. This has the effect of diverting our attention away from previous cases where the method has failed and hence of escaping the sceptical argument which takes its start from those cases. But we might reasonably doubt whether the requirement that the method be reliable this time amounts to any genuine addition to the tripartite account. If reliability is defined in terms of the production of truth, it adds nothing to the first condition once we restrict our attention to the particular case. If it is defined in terms of justification, it adds nothing to the third. And no other accounts seem very inviting. (It may be, however, that the causal theory amounts to a notion of justification in the particular case; see 2.4)

Conclusive Reasons

A different approach diagnoses Henry's failing in the Gettier case as due to the fact that his reasons were less than conclusive. If we require for knowledge that the justified true belief be based on conclusive reasons, all the Gettier cases, and indeed any case in which the believer is right by accident, fall to the ground.

All the work in this approach must go into a persuasive account of what it is for reasons to be conclusive. One suggestion would be that where beliefs A – M constitute conclusive reasons for belief N, A – M could not be true if N is false. This will exclude the counter-examples, but it will also make knowledge a rare phenomenon at best. Empirical knowledge, anyway, looks impossible now; in the empirical realm, our reasons are never conclusive in this sense.

A weaker account, owed to F. Dretske (1971), suggests that someone's reasons A − M for a belief N are conclusive iff A − M would not be true if N is false. This is weaker because to say that A − M *would* not be true if N is false is not to say that they *could* not be true if N is false, as the stronger account demands. It is so weak as not really to provide a genuine sense of 'conclusive', but this doesn't really matter. This weaker account seems to me promising in its general approach, and the theory I shall be supporting in the next chapter is distinctly similar. But it differs in not talking about reasons; and this is a virtue because it does seem possible that there should be justified belief without reasons. My belief that I am in pain may be justified, perhaps, but I can hardly be said to base it on reasons, conclusive or otherwise. I don't base it on reasons at all.

The Causal Theory

A. I. Goldman proposes a causal supplement to the tripartite definition (Goldman, 1967). An initial diagnosis of the Gettier counter-examples may be that it is just luck that Henry's justified belief is true. This diagnosis cannot itself provide a suitable answer. We cannot merely stipulate that there be no luck involved, because we all of us rely on luck to some extent. For instance, the fact that our reliable belief-gathering method provides here a true belief rather than a false one, as it sometimes does, will be just luck as far as we are concerned. And of course the fact that luck is always involved somewhere gives the sceptic a toehold too. But the diagnosis can suggest a better answer. Goldman's suggestion is that what made the belief true in the Gettier case is not what caused Henry to believe it. So he proposes, as a fourth condition for knowledge that p, that the fact that p should cause a's belief that p. This excludes the Gettier cases because in them it is coincidental that the belief is true. We want a link between belief and truth to prevent this happening, and a causal link looks promising.

Attractive though this approach is, it faces difficulties. The first is that we may find it hard to suppose that facts can cause anything; surely they are too inert to affect the way the world goes, even where that world is the merely mental world of beliefs. What, after all, are facts? One's first idea is that facts are similar to, if not identical with, true propositions (which would explain why there are no false facts). But can true propositions cause anything? Surely facts (or true propositions) reflect the world rather than affect it. The prevalent analyses of causation seem justifiably only to allow events

and possibly agents as causes. Second, there is a problem about knowledge of the future; Goldman's suggestion seems to require that here either we have an instance of backward causation (the future causing the past) or that knowledge of the future is impossible since causes cannot succeed their effects. Third, there is the problem of universal knowledge, or more generally of knowledge by inference. My belief that all men are mortal is caused, but not by the fact that all men are mortal; if any facts cause it, they are the facts that this man, that man, etc., have died. And these men are not caused to die by the fact that all men die (which would restore the causal analysis, with an intermediary cause); rather, all men die because those men do (among others). How then can the causal analysis show that I know that all men die?

There are answers to some of these criticisms, of course. We are more used to talking about facts as causes than the first criticism allows. The fact that philosophers have not yet persuaded themselves that they understand the idea that facts can be causes should not cause us to rule out all appeal to fact-causation as philosophically unsound. (The preceding sentence is a case in point.) The second criticism, too, might be answered by complicating the theory by allowing facts to be known in cases where fact and belief are different effects of a common cause. The third criticism, however, seems more intractable. The admission that facts can be causes will not much improve our willingness to suppose that universal facts can cause universal beliefs.

There are promising aspects about the causal theory, and the theory which I shall support can in fact be seen as a generalization from it.

2.4 CONCLUDING REMARKS

The various proposals considered in the previous section were presented as if they were additions to the tripartite analysis, it being admitted that Gettier had shown that analysis to be insufficient. But we can find among them at least one which can be seen as a direct defence of the tripartite analysis. Any proposal which amounts to a new theory of justification may succeed in showing that in the Gettier cases the relevant true beliefs were not justified at all. And we could take the causal theory in this way. The causal theory could be telling us that a belief is only justified when caused (directly or indirectly) by the facts. It would then be adopting route 1, as

distinguished in 2.2. (Some versions of the reliability proposal could also be seen in this light.) Moving this way, then, we would be starting from a causal theory of justification; the causal theory of knowledge would simply be one of its consequences.

A possible way of arguing against a causal theory of justification would be to claim that we have no guarantee that there is only one way in which beliefs come to be justified, and in particular no real reason for supposing that any acceptable way must somehow be causal, so that all justified beliefs that p must be caused by relevant facts. Surely we don't want to rule out in advance the possibility that some moral beliefs, say, are justified, doing so just because we don't want to admit the existence of moral facts (if we don't). And we might still be doubtful of the existence of causally effective mathematical facts, without wishing to say that no mathematical beliefs can therefore be justified.

More importantly, however, the suggested causal account of justification is false because it denies the possibility that a false belief be justified. A false belief that p has no fact that p to cause it. This objection can only be evaded by finding a different account of the justification of false beliefs from that which is offered for true ones. But that cannot be right. Justification must be the same for true as for false beliefs, if only because we can ask and decide whether a belief is justified (e.g. a belief about the future) before we decide whether it is true or false.

This criticism leaves open the possibility of a different sort of causal theory, on the lines suggested at the end of 2.2. With a causal theory of knowledge and the thesis that a belief is justified iff if true it would be knowledge, we can give a causal account of justification which is not vulnerable to the existence of false justified beliefs.

FURTHER READING

Central papers in the area are Gettier (1963), Dretske (1971), Goldman (1967) and Swain (1974).

Perhaps the earliest discussion (and rejection) of the tripartite definition is in Plato's *Theaetetus* (Plato, 1973, 201c – 210d).

The enormous industry recently generated by perceived defects in the tripartite account is painstakingly analysed in Shope (1983), with copious references. There are of course many approaches and variants on approaches to the Gettier problem which I have not discussed, including Shope's own.

Most of the papers referred to in the present chapter are collected in Pappas and Swain (1978), which also contains an analytical introduction to the area.

Prichard (1967) gives an interestingly different account of the relations between knowledge, belief, certainty and truth.

An important question which we have not discussed is whether knowledge implies belief. For this, cf. Ring (1977).

The papers by Gettier, Prichard and Woozley are collected in Phillips Griffiths (1967).

3

The Conditional Theory
of Knowledge

3.1 THE THEORY

This theory, which we owe mainly to Robert Nozick, takes its start as others do: from the defects which Gettier exposed in the tripartite analysis. Nozick suggests that the reason why we take the justified true beliefs in those examples not to have been known is that *a* would have believed them even if they had been false. The reason why Henry's belief that McEnroe is this year's champion was too lucky or too luckily true to count as knowledge is that his route to this lucky truth was such that even if it had been false, he would still have ended up believing it. Nozick takes it therefore that for *a* to know that *p* we require that *a* would not have believed that *p* if *p* had been false.

This gives us, so far, the standard two conditions:

1 *p*
2 *a* believes that *p*,

with

3 if *p* were not true, *a* would not believe that *p*.

But Nozick argues that although this account may cope with the examples Gettier offers, there are other similar examples which would escape what we have so far. There are two ways in which it can be a coincidence that *a*'s belief is true, and both need to be ruled out. The first is that if it were false, *a* would still believe it; we have dealt with this already by the addition of clause 3. The second is that there may be slightly different circumstances in which it remains true, but *a* no longer believes it. Many examples are

coincidental (lucky) in both senses. Suppose that I believe that there is a police car in the road outside because I can hear a police siren. There is in fact such a car outside, but the siren I hear is on my son's hi-fi in the next room. I do not know that there is a police car outside, for two reasons. First, I would still have had the belief even if the car had not been there. Second, I would not have believed the car was there if my son's hi-fi had been silent, even though the car itself remained outside. We rule out this second way in which a true belief may be too luckily true to count as knowledge by adding a fourth clause to the initial three:

4   if, in changed circumstances, p were still true, a would still believe that p.

These four conditions, comprising the conditional theory of knowledge, can be symbolised as follows:

1   p
2   Bap
3    $\sim p \; \Box\!\!\rightarrow \; \sim$B$ap$
4    $p \; \Box\!\!\rightarrow \;Bap$

where the bow-arrow '$\Box\!\!\rightarrow$' is used to symbolise the subjunctive conditional construction in English 'If it were the case that. . . , it would be the case that. . . '. The theory which they comprise is an attempt to articulate the feeling that, for a belief to be knowledge, it must be peculiarly sensitive to the truth of the proposition believed; it must *track* the truth (Nozick's term) in the sense that if the proposition were in changed circumstances still true, it would still be believed, and if it were not true, it would not. In the case of the police car, my belief fails to track the truth in both ways, and so is not knowledge.

3.2 SOME COMMENTS

(a) The requirement that the belief that p should track the truth of p is a requirement that the first two clauses of the theory should be related in a certain way. This is similar to the way in which the causal theory worked, but in that case the relation required was specifically a causal relation. The conditional theory is less demanding, and hence escapes some of the difficulties which the causal

theory faced. But it includes the causal theory as a special case, since we might think that if the fact that p does cause a's belief that p, then the two subjunctive conditionals will be true (but not vice versa). So the conditional theory is a generalisation of the causal theory. While the causal theory suggests that there is only one way in which justified true beliefs can get to be knowledge (by being caused by the facts), the conditional theory is willing to countenance any way, causal or not, which preserves the truth of the two subjunctive conditionals.

In fact the conditional theory adopts some of the best points of several of the theories found wanting in the previous chapter. For instance, it is close to Dretske's version of the 'conclusive reasons' approach.

(b) We considered whether the causal theory of knowledge either rested on or made available a causal theory of justification. So we should ask the same questions of the conditional theory. Can we offer a conditional definition of JBap, thus:

$$\text{JB}ap \equiv (p \; \Box\!\!\rightarrow \text{B}ap \; \& \; \sim p\Box\!\!\rightarrow \; \sim\text{B}ap)?$$

This would be to hold that a justified belief is one which tracks the truth. But the same sort of difficulty arises. A false belief may nevertheless be justified. If so, JBap is consistent with (Bap & $\sim p$). But (Bap & $\sim p$) is inconsistent with ($p \; \Box\!\!\rightarrow$ Bap & $\sim p \; \Box\!\!\rightarrow \sim$B$ap$); a false belief does not track the truth. Hence the conditional definition of justification fails.

Such a theory would have been attractive, however. For someone who takes his belief that p to be justified is surely close to taking it that his belief tracks the truth of p. We can, if we want, restore the possibility of a conditional theory of justification by a move like the one made in 2.4. The suggestion there was that we derive the account of justification from that of knowledge, thus: a is justified in believing p iff in certain circumstances a would know that p. As an illustration, we can offer this instance of that approach:

$$\text{JB}ap \equiv (p \; \Box\!\!\rightarrow \text{K}ap).$$

Here the crucial phrase 'in certain circumstances' is read in the simplest possible way, as 'if p were true'. If a theory like this were available, we would have reintroduced in a roundabout way the possibility of a conditional theory of justification.

(c) Another point which I take to be in favour of the theory is that

it begins to make some theoretical sense of the intuitive feeling that what was wrong in the Gettier cases was that there was too much luck around. The theory gives an account of what it is for a belief to be luckily true, as follows: the extent to which a's belief is luckily true is the extent to which even if it had been false, a would still have believed it, or if it were in changed circumstances still true, he would still believe it. The importance of this account will emerge in the next section.

(d) The theory does seem to have some resources with which to explain the link between certainty and knowledge. Someone who claims that he knows that p is claiming that if p were not true, he would not believe it and if p were true he would believe it. But this claim is precisely one which he would not make if he were not certain that p. What the diffident schoolboy has lost is the confidence that his beliefs are tracking the truth; he takes it that, although he does believe that p, it is at least as probable that he is wrong as that he is right. The theory therefore analyses the certainty required for a knowledge claim as the belief that the two subjunctive conditionals are satisfied. It uses this analysis to explain the otherwise puzzling fact that the diffident schoolboy does know but cannot claim to know.

3.3 THE PRINCIPLE OF CLOSURE AND THE FIRST SCEPTICAL ARGUMENT

The conditional theory of knowledge can show that you do not know that you are not a brain in a vat. For a necessary condition for such knowledge, 3 (that if it were not true that you are not a brain in a vat, you would not believe that you are not a brain in a vat) fails. We can simplify this conditional as follows: if you were a brain in a vat, you would not believe that you are not a brain in a vat. This is false because if you were a brain in a vat (being fed your current experiences), you would still believe that you are not a brain in a vat. Therefore, if the conditional theory of knowledge is on the right lines, you do not know that you are not a brain in a vat. A necessary condition for such knowledge is not satisfied.

Despite this, Nozick's account can be used to show that you do know that you are currently sitting reading a philosophy book (please sit down first). For the four conditions are all satisfied. It is true that you are sitting etc., you believe that you are sitting etc., if you were not sitting etc. you would not believe that you are, and

if you were sitting etc. you would believe that you are.

The account also shows, in similar fashion, that you know that if you are sitting reading a philosophy book, you are not a brain a vat. And so it emerges that on the conditional theory of knowledge it is possible to know that p and to know that p implies q without knowing that q. But doesn't this conclusion directly breach the principle of closure $[Kap \& Ka(p \rightarrow q)] \rightarrow Kaq$? It does; but Nozick is able to show more generally that on his account that principle fails, and (which is perhaps more important) to explain its failure. The explanation relies on the broad sweep, but not the details, of a theory about the conditions under which subjunctive conditionals such as 3 and 4 are true.

The theory uses the notion of a possible world in order to give its account of truth conditions for subjunctive conditionals. A possible world is to be thought of as a complete way in which the world might have been. The thought that the world might have been different in a certain respect is taken to be the thought that there is a possible world which does differ from the actual world in that respect (and probably others too). Possible worlds vary in their degree of resemblance to the actual world. Some are close to our world, others much more remote. But it may not be possible to order worlds according to their closeness to the actual world, for two reasons. First, the notions of closeness, resemblance and similarity are too imprecise to support the sort of precise comparative judgement which such an ordering would require. Second, for any possible world we can reasonably expect to be able to find another which resembles the actual world to the same degree. For both these reasons we need to think not in terms of individual possible worlds but of groups of possible worlds, all worlds in a group being equally close to or distant from the actual one. If there is even a vague or sketchy ordering, it will be an ordering of groups of equally similar worlds rather than of individual worlds.

Nozick gives an initial account of truth-conditions for a subjunctive conditional $p \;\square\!\!\rightarrow q$ as follows:

$p \;\square\!\!\rightarrow q$ is true in the actual world iff $p \rightarrow q$ is true throughout a range of groups of worlds reasonably close to the actual world.

(This is not the final theory but a first approximation.)

Although this theory is expressed formally it can easily be given intuitive support. Suppose that we want to know the circumstances which would justify assertion of the subjunctive conditional 'If

Mrs Thatcher had waited longer, she would have lost the election'. What we do is to picture the current situation altered to the extent that Mrs Thatcher delays the election. Other things will have to be altered as well, of course, for the date of the election is not an isolated fact (there are no isolated facts, which is why there cannot be a world similar to ours in every respect except one). We will have also to change all those things which would not have happened if the election had been delayed, e.g. Mrs Thatcher's age at election time. Then, holding everything else constant so far as we can, we form a judgement about what will most probably happen. If we take it that a Conservative defeat is most probable, we assert the subjunctive conditional. If not, not.

The formal theory implements this informal approach by saying that we consider the nearest groups of worlds in which the antecedent p is true and ask whether in those worlds q is true also. If q is probable given p, then p & q is more likely than p & $\sim q$; and therefore in the nearest worlds where p is true, q will also be true. There will of course be some remoter worlds where we have p & $\sim q$, but they do not matter. For we are deciding what would (most probably) happen, not what might (within the bounds of possibility) happen.

This distinction between what would happen and what might happen is crucial for what follows. To take David Lewis' example (D. Lewis 1973), 'if kangaroos had no tails, they would topple over'; it is of course always true that they *might* not topple over − they might be given crutches by a grateful and tourist-conscious government. But this doesn't affect the fact that they would topple over, for in worlds most similar to ours tailless kangaroos do topple over and are not given crutches.

Nozick's account can now be used in two ways. First, it can reinforce our intuitions that you know that you are sitting reading etc., that you know that if you are sitting reading etc. you are not a brain in a vat, and that you do not know that you are not a brain in a vat. With the conditional theory of knowledge, and the above explication of clauses 3 and 4, you can be said to know that you are sitting reading because (3) in the nearest worlds in which you are not sitting reading you do not believe that you are, and (4) in the nearest worlds in which you are sitting reading you do believe that you are. However, you do not know that you are not a brain in a vat because it is not the case that (3) in the nearest worlds in which you are a brain in a vat you believe that you are a brain in a vat. I emphasise again that it is only the *nearest* worlds in which the

antecedent is true that count. It is admittedly possible, for instance, that you should believe you are sitting reading when in fact you are not. But such a world is vastly more remote from the actual world than is a world in which, when you are sitting reading, you believe that you are. We are interested not in what *might* be the case (of course anything is possible) but in what *would* be the case, and the theory captures that interest by focusing only on the nearest relevant worlds.

The theory also provides a direct disproof of the principle of closure. This can be done in two ways, either by giving a description of a world for which the principle fails and and proving formally that the description is consistent, or informally by appeal to a counter-example. We already have such an example: the case where a = you, p = you are sitting reading, and q = you are not a brain in a vat. But this example is contentious. It needs support from an explanation, in terms of the conditional theory, of how there can be such counter-examples. This can now be done.

An important preliminary task is to distinguish the principle of closure, which is false, from the following much more reliable instance of *modus ponens*:

$$[Kap \ \& \ (Kap \rightarrow Kaq)] \rightarrow Kaq.$$

So long as these two principles are kept apart, it should be easy to see why the principle of closure fails. The reason lies in the subjunctive conditionals and the possible worlds held relevant to the assessment of those conditionals as true or false. The worlds relevant to assessing the left-hand side of the principle of closure are the nearest group in which p is true, the nearest group in which p is false, the nearest group in which $p \rightarrow q$ is true (which may be distinct) and the nearest group in which $p \rightarrow q$ is false. The worlds relevant to assessing the right-hand side of the principle of closure are the nearest group in which q is true, and the nearest group in which q is false, i.e. completely different sets of worlds one of which may be, as in our actual example, far more remote from our world than are any of the first four groups. Hence it is hardly probable that a group of remarks about the first four groups could place any restriction on the nature of the two later groups, and thus there is a lot of room for an invalidating counter-example to the principle of closure. The more distant one of those later groups is, the more likely we are to be able to construct a counter-example; as is revealed in our own example where $\sim q$ = you are a brain in a vat.

By way of final illustration, we can apply Nozick's refutation to Descartes' sceptical argument about dreaming, and hence perhaps do better than Descartes did on the last page of the *Meditations*. Here a = you, p = you are sitting reading, q = you are not in bed dreaming that you are sitting reading. Can we have Kap, $\sim Kaq$ and $Ka(p \rightarrow q)$? Yes. Kaq is false because in the nearest group of worlds in which q is false (you are in bed dreaming etc.) you still believe that q is true. Kap is true because in the nearest worlds in which you are sitting reading you believe that you are sitting reading, and the nearest worlds in which you are not sitting reading are either worlds in which you are standing/kneeling/lying reading or worlds in which you are sitting knitting/watching TV, etc., but not worlds in which you happen to be asleep dreaming that you are sitting reading. Hence you can know that you are sitting reading even though you do not know that you are not asleep in bed, dreaming that you are sitting reading.

3.4 HAS NOZICK REFUTED THE SCEPTIC?

One purpose of the preceding section was to show that an interest in the strengths and weaknesses of the tripartite condition can yield genuine philosophical profit. The conditional theory is attractive ir :s own right as a promising account of knowledge which escapes Gettier-type objections. And it has the secondary virtue that it destroys a prevalent type of sceptical move, and does so in a way that explains the attraction of that move.

According to Nozick, however, all sceptical arguments rely on the principle of closure and the invalidation of that principle serves therefore as a general response to scepticism (Nozick, 1981, p. 204). But apart from the general implausibility of this suggestion, there is surely one sceptical argument which he cannot reject in this way. This is the argument from error (1.2).

The argument is that we or others have made mistakes in the past or would make them in circumstances which, so far as we can tell, are not relevantly different from our present circumstances. We cannot therefore, admitting that we or they did not know before, insist that we do know now, since that would be to make different claims in circumstances that show no relevant difference.

This argument turned out to be the one that forced us to admit in the first place that we do not know that we are not brains in vats. Nozick needs this conclusion to generate the sceptical problem that

he is attempting to dissolve. For his sceptic argues that the consequent Kaq of PC^k is here false, and hence, via the admitted truth that $Ka(p \rightarrow q)$, that Kap is false. Nozick wants to unroll this sceptical argument while admitting its first move, and it is the argument from error that forces that admission. So it is only if he admits the strength of the argument from error that he can think he is getting anywhere by refuting PC^k.

Nozick might claim that his conditional theory also proves the same point, that we don't know we are not brains in vats. And he is right; our belief here would not track the truth. But if we had no independent reason for accepting this conclusion, such as the argument from error provides, we would take it as a point *against* his theory that it shows we don't know the most central and obvious things such as that we are not brains in vats, that there is a material world or that the world began more than five minutes ago. So he cannot cast off the argument from error and rely entirely on the conditional theory to show that Kaq is here false.

Second, there is something right about an argument which holds, as does the argument from error, that for the very same sort of reason that you don't know you are not a brain in a vat, you don't know, e.g., that *The Times* will be published tomorrow, nor whether you are sitting reading. Surely it is implausible to suppose, taking Russell's example, that I do know what I did yesterday but I don't know that the world did not begin five minutes ago replete with archaeological and other traces (Russell, 1921, p. 159). Nozick wants to show us that we can suppose this; but I am not sure that I really want to (or ought to want to). The argument from error has here a plausible consistency, while the point which Nozick takes to be his strength begins to look like a weakness. (This remark could be promoted to form the basis of a general argument against the conditional theory of knowledge.)

Third, the argument from error cannot be attacked for relying on PC^k. This is not because PC^k is valid, particularly. Even if it were valid, I can see no reason for supposing that the argument relies upon it. Nozick anyway could not hold that it does, because his position relies on the independent proof it provides that you don't know you are not a brain in a vat. Without that proof, we would hold it against the conditional theory that it cannot even show that we know that we are not brains in a vat. With it, Nozick can and does find independent support for his theory; he wants to say that the theory gets things right here. But if the argument from error does rely on PC^k, which on his theory is invalid, he lacks

that independent support and with it a considerable degree of plausibility. And I do not know of any other argument, not relying on PC^k, which Nozick could use to show that we do not know we are not brains in a vat without showing that we do not know most other things either.

Fourth, the argument cannot be rejected on the similar grounds that it argues directly from the fact that you might be wrong to the conclusion that you do not know. Such an argument would be invalid, as Nozick can show. But the argument does not proceed directly in this fallacious way, but indirectly via the principle of universalizability.

3.5 INTERNALISM AND EXTERNALISM

So far we have it that Nozick's account of knowledge succeeds in defusing one sceptical argument but not another. Is this a conclusion which Nozick must simply accept, abandoning his claim to total success and pointing merely to the admitted partial success?

There remains a complaint which Nozick could and probably would make. He could say that his conception of knowledge is an *externalist* conception, while our sceptical argument from error is an *internalist* one. If externalism is a sound stance, then the argument from error is irrelevant; for it does no more than elaborate on a defective (though traditional) approach to epistemology or if it does more, it succeeds only in showing how that defective approach must lead to scepticism.

What is meant by an 'externalist' and an 'internalist' conception here? The answer to this question can best be given by example. The causal theory of knowledge, which defines 'a knows that p' as equivalent to

1   p is true
2   a believes that p
3   a's belief 2 is justified
4   a's belief 2 is caused by the fact that p is true

is an externalist conception, because condition 4 is one which a might be entirely incapable of recognizing or pointing to when asked whether he does know that p. The externalist says, in this case, that so long as condition 4 does in fact hold, whether a is able to point it out or even to understand it or not, a does know that p (given

conditions 1 – 3, of course). The internalist would claim that for the causal clause to turn justified true belief into knowledge, it must not only be true but be believed by *a* to be true. Thus the internalist would add:

5   *a* believes 4.

There are arguments in favour of externalism and arguments in favour of internalism. The externalist can point out how difficult it is going to be for the internalist to provide a satisfying account of knowledge. Surely, he might say, if we are to add clause 5 we should also add

6   *a*'s belief 5 is justified,

and then presumably

7   *a*'s belief 5 is caused by the fact that 5 is true.

But we have surely now generated an infinite regress, which will mean that internalism is doomed to scepticism. What is more, the regress does not depend on the causal element of the example used. We could create the same regress by starting from the traditional tripartite conception and adding, on internalist grounds,

4   *a* believes 3

and then presumably requiring

5   *a*'s belief 4 is justified

and then

6   *a* believes 5

and so on.

The internalist can respond by pointing out how great our natural intuition is to favour the internalist conception. Suppose that we work with the causal example; it is required for knowledge that the fourth clause be true, but not that *a* have any inkling that the fourth clause be true. Doesn't this show that, for all *a* knows, he doesn't know that *p*? And how can he know that *p* when for all he knows he doesn't?

In my opinion neither of these arguments is effective in destroying its opponent. The first merely points out the difficulties with

scepticism; the internalist would accept this and say that these difficulties have to be faced and not ignored. The second seems to amount more to a statement of the internalist position than an independent argument against externalism. In fact I doubt that there can be a conclusive argument in favour of either of these approaches; the approaches are so different that there is a danger that any argument will simply beg the question.

Now Nozick's position is intended to be externalist (Nozick, 1981, pp. 265 – 8 and 280 – 3). Conditions 3 and 4 are obvious evidence of externalism, and there is no suggestion that we need to add anything like

5   $Ba(\sim p \ \Box \rightarrow \ \sim Bap)$.

As an externalist, would he not be justified in simply dismissing the argument from error as an irrelevant expression of internalism? I do not think so, for two reasons. First, again, Nozick relies on the argument, internalist or not, to provide the independent support for something which would otherwise look like a counter-intuitive consequence of his theory. Second, in taking the argument to be necessary Nozick tacitly accepts that his theory is not as purely externalist as it might initially appear, and hence he cannot claim immunity from an attack merely on the grounds that it is internalist.

We will return to the contrast between internalists and externalists in 9.3, and examine it in greater detail. Meanwhile, where has our discussion of the conditional theory left us? We still have no answer to the sceptical argument from error. We have mentioned, but not adopted, two positions which show promise in this respect: externalism and anti-realism. But those who find these positions unattractive will have to look further if they want to say that there are such things as knowledge or justified belief. The argument from error will recur as a persistent threat. I shall venture my own response to it in the final chapter.

FURTHER READING

The conditional theory and its consequences for scepticism are expounded in Nozick (1981) pt 2, especially pp. 172 – 8 and 197 – 227.

Stroud (1984, ch. 2) considers a similar argument against the sceptic's attempt to generalize from the fact that we do not know we are not brains in a vat. He continues the discussion in his ch. 7.

Counter-examples to the conditional theory are given in Garrett (1983); Gordon (1984) attempts an answer. Shope (1984) offers more than counter-examples.

D. Lewis (1973) gives a pioneering account of truth-conditions for subjunctive conditionals (which he calls "counterfactuals") in terms of possible worlds. I discuss an important difference between Nozick and D. Lewis in Dancy (1984b).

References to the internalism/externalism debate are given at the end of ch. 9.

Part II
JUSTIFICATION

4
Foundationalism

4.1 CLASSICAL FOUNDATIONALISM

Perhaps the most influential position in epistemology is the one I shall call classical foundationalism. Discussion of justification, of what it is for a belief to be justified, begins with this theory; other theories will be described in terms of their relation to or divergence from this one. It offers a compelling picture of what the aims of epistemology are; in short it amounts to a definition of the epistemological enterprise.

The classical foundationalist divides our beliefs into two groups: those which need support from others and those which can support others and need no support themselves. The latter constitute our epistemological foundations, the former the superstructure built on those foundations.

This distinction between foundations and superstructure, between basic and non-basic beliefs, is a structural one. But classical foundationalism gives the distinction content by adding that our basic beliefs are beliefs which concern the nature of our own sensory states, our own immediate experience. Such beliefs are able to stand on their own feet, without support from others. Other beliefs need support, and hence must get it from our beliefs about our sensory states.

Classical foundationalism thus gives expression to the central tenet of empiricism, the view that all our knowledge is derived from our experience. It does this by insisting that a belief which is not about our own sensory states (immediate experience) must, if it is to be justified, be justified by appeal to beliefs which are about our own sensory states.

How is it that beliefs about our present sensory states need no support from others, while all other beliefs require such support? The answer comes from the third element of classical foundationalism;

this is that our beliefs about our present sensory states are infallible. It is because of this that they can play the role ascribed to them in this form of empiricism; beliefs about our present sensory states can be our basis — can stand on their own two feet and support the rest — because they are infallible.

We can now see what epistemology is, according to classical foundationalism. It is a research programme which sets out to show how it is that our beliefs about an external world, about science, about a past and a future, about other minds, etc., can be justified on a base which is restricted to infallible beliefs about our sensory states. It is suggested that if we can do this, the demands of epistemology are satisfied. If not, we relapse into scepticism.

In this chapter and the next we shall examine classical foundationalism in some detail, and find reason to reject almost every part of it. But first we should turn to investigate the motives and arguments which lead or have led philosophers in this direction. We have already seen that classical foundationalism is an expression of empiricism. But there are other expressions of empiricism, as we shall see. Why should we opt for this one?

Probability and Certainty

C. I. Lewis, the most eminent classical foundationalist of this century, held that "unless something is certain, nothing else is even probable" (see C. I. Lewis, 1952). This view can be best understood by approaching it from a (very slight) knowledge of the probability calculus. In this calculus, probability is always assessed relative to evidence. We do not ask what the absolute probability of a hypothesis h (written $P(h)$) would be. Instead we ask about h's conditional probability given evidence e (written $P(h/e)$). The probability of h given e is expressed as correlations generally are, on a scale from 0 to 1. If $P(h/e) = 0$, then given e, h is certainly false. If $P(h/e) = 1$, then given e, h is certainly true. If $P(h/e) = .5$, then it is as probable that h is true, given e, as that it is false, since $P(h/e) + P(\sim h/e) = 1$ in the calculus.

The main point is that in assessing the probability of h given e we do not question e; we assume temporarily that e is certain, and we ignore the chance of e not being true. But e itself has a probability relative to further evidence e', and so on indefinitely. And unless we can find in the end a proposition or set of evidence e^n which has somehow in its own right the probability 1, all these probabilities will have nothing to rest on. We need to find something certain which can function as the unquestioned evidence by appeal.

to which the probabilities of other things are to be assessed.

In this argument it is suggested that a proposition with probability 1 is *certain*. But certainty and infallibility are not identical, and we are trying to explain a theory which takes its basic beliefs to be infallible. The move from one to the other is easy, however. If a proposition, being certain, has a probability of 1, then there is no chance that a belief in that proposition will be false; so the belief will be infallible.

There is an oddity in the argument, which starts by insisting that we speak only of probability relative to evidence, and ends by talking of a proposition having a probability of 1 in its own right. Probability theorists escape this oddity by defining absolute probability in terms of relative probability: they say that the absolute probability of h = the probability of h relative to a tautology. ($P(h) = P(h/q v \sim q)$.) Whether this manoeuvre is more than a technical device is dubious.

The Regress Argument

All agree that some of our beliefs are justified by their relation to other beliefs. Standardly that relation is thought of as inferential; one belief is inferred from another or others. Thus my belief that striking the match will light it is justified inferentially. I have inferred it (not consciously, of course) from other beliefs, probably beliefs about similar occasions in the past.

The regress argument is an argument that as well as the inferentially justified beliefs, there must be some beliefs which are justified non-inferentially. The general thrust can be seen intuitively by supposing that inference is basically a matter of moving from premises to conclusion along an acceptable path. If the premises are unjustified, there will be no justification for the conclusion – at least, not by *this* inference. We can suppose, then, that only justified beliefs can justify others; and it is this thought which generates the regress.

Suppose that all justification is inferential. When we justify belief A by appeal to beliefs B and C, we have not yet shown A to be justified. We have only shown that it is justified if B and C are. Justification by inference is conditional justification only; A's justification is conditional upon the justification of B and C. But if all inferential justification is conditional in this sense, then nothing can be shown to be actually, non-conditionally justified. For each belief whose justification we attempt there will always be a further belief upon whose justification that of the first depends, and since this regress is infinite no belief will ever be more than conditionally justified.

The only way round this conclusion is to suppose that instead of the chain of justification stretching away to infinity, it turns round on its tail and joins up with itself at some point, forming a circle. But this will not mend matters, for it will still be the case that the justification of all members of the loop is conditional. The loop will never succeed in removing the conditionality.

The regress argument therefore drives us to suppose that there must be some justification which is non-inferential if we are to avoid the sceptical consequence of admitting that no beliefs are ever actually justified. And the claim that there are two forms of justification, inferential and non-inferential, is the core of any form of foundationalism in the theory of justification.

There is a variety of possible responses to the regress argument, apart from straight capitulation. A central one will be suggested in 9.1. Meanwhile we should ask if the regress is as damaging as it might seem. Not all infinite regresses are vicious. Some are *virtuous*, i.e. we can live with them and do not have to find some way to stop them. For instance, the regress generated by the remark that there is a point between every pair of points may be virtuous, even when we take it to concern points in time rather than in space. Equally we might accept the temporal regress caused by supposing that for each moment in time there is a moment that precedes it, or the causal regress derived from the propositions that every event has a separate cause and that every cause is an event. We might even accept the regress caused by the suggestion that when we believe that p we believe that p is probable (the regress comes by taking q = 'p is probable'). Can we not then simply accept that justification continues *ad infinitum*? I think that the regress of justification, once it has been allowed to start in the way I have outlined, is vicious in the sense that it will show that no belief is ever actually justified. There is a bad reason for thinking this, which involves taking the regress as temporal; *before* I can justify A, I have to justify B and C, and so on *ad infinitum*, and thus I can never get started. But I do not take the regress argument to be concerned with temporal relations between acts of justification. A better approach merely underlines what was said above, that the regress shows that if all justification is inferential, no belief is ever more than conditionally justified. If knowledge requires more than conditional justification, as it seems to, then the only way to escape the sceptical thrust of the regress argument is to conclude with the foundationalist that some beliefs are justified non-inferentially.

If we are to avoid this foundationalist conclusion we shall have

to show that the regress argument is fallacious. I shall give later (in 9.1) a non-foundationalist answer to it. At present I only wish to draw attention to a possible ambiguity in a crucial move. The sentence 'We have only shown that A is justified if B and C are' could mean, as was pretended above, that we have shown A's justification to be conditional on that of B and C; but it could also mean that if B and C are in fact justified, we have shown that A is, i.e. that the success of our demonstration is conditional, not the justification we have demonstrated. The argument as I have presented it seems to require the first reading of this crucial sentence; on the second reading we get, not a regress of justification, but a demonstration of justification which is only successful in certain conditions.

Our regress argument differs from Lewis' argument about probability and certainty, despite the great structural similarities (they are both regress arguments, really). The difference between them is that the first regress can only be stopped by certain (= infallible) beliefs, while the second insists merely on the existence of beliefs which are non-inferentially justified.

Infallibility and Justification

The two previous arguments are brought together as a joint argument for classical foundationalism by the claim that any infallible belief would be non-inferentially justified. An infallible belief would be justified but would not derive its justification from any relation in which it stood to other beliefs; it would not need any support from elsewhere. For surely a belief whose chances of being false are nil is unimpeachable. Nothing could reduce its probability, and hence there could be no reasons for supposing it false. So if there are any infallible beliefs we have no need to worry about the threatened regress of justification. Infallibility in the base will stop the regress.

We shall see in 4.3 that even if all infallible beliefs are non-inferentially justified, the reverse is not true; this is what opens the door for foundationalisms other than the classical variety. One can abandon the view that we have any infallible beliefs and find other ways of supposing that some beliefs are non-inferentially justified. But we cannot just announce that this is the case, e.g. with our beliefs about our sensory states. We have to produce some account of how it is that a belief can achieve this status and play this special role. Classical foundationalism says, plausibly, that our beliefs about our sensory states can do this because they are infallible. In the next section I offer an argument that this cannot be right, and if the

argument is sound we shall have to find some other way of showing how any of our beliefs can be non-inferentially justified and thus can stand on their own feet.

4.2 PROBLEMS FOR CLASSICAL FOUNDATIONALISM

One of the main reasons for wanting one's own basic beliefs to be infallible is that this would guarantee that they are all true. But is there any real purpose in seeking this guarantee? The principles of inference by which we are to move from basic to non-basic beliefs are fallible, in the sense that they take us sometimes from true beliefs to false ones. (Remember Russell's chicken (Russell, 1959, p. 35), whose true beliefs about the regularity with which it had been fed so far led it into a false belief about the security of its future.) If there is this source of contamination necessarily present in the procedure, why should we insist that the input to the procedure be completely sterile, i.e. devoid of any taint of falsehood?

But the main objection to classical foundationalism is that there are no infallible beliefs. The *fallibilist* holds, correctly in my opinion, that we are nowhere entirely immune from the possibility of error.

Are our beliefs about our present sensory states infallible? Champions of infallibility tend to concede that there is room for a mistake in the *description* of one's sensory states (see Ayer, 1950). I might mistakenly describe my sensory state as being an experience of pink (things look pink to me here) when in fact it is an experience of orange. But this is dismissed as a merely verbal error. Of course I can be mistaken about the meanings of the words I use, but this will not show that I have any mistaken beliefs about my present sensory states. On the contrary, I must know how things look to me: my only error lies in choosing the wrong words to describe it. The description I use may be false, but I, the describer, am in this case infallible. My *beliefs* − the things I use words to express, with more or less success − must be true.

Similarly, we may say that merely verbal errors can be corrected in standard ways. You can show me or remind me of the difference between orange and pink, perhaps by showing me a colour chart. When I have grasped this (the difference, not the colour chart), I can apply it to my present experience in order to see whether the experience is one of pink or of orange. But in order to do this I must be aware of the nature of the experience already, before I

compare it with others in order to get the right words to describe it. I don't change my beliefs about how things look, only about how to describe them.

Third, although some comparison between my present experience and others is necessary for me to know what words to use in description, and although such comparison, especially in the case where the objects compared are a past and a present experience, is fallible (since memory is fallible), still the comparison is not what I am trying to express when I try to express my beliefs about my present experience alone. For my experience would have been this way no matter how other experiences might or might not have been. So the fallibility of the comparison does not extend to show the fallibility of the belief expressed; it only shows the fallibility of the expression of belief.

Finally, if a comparison is possible at all, this can only be because ultimately we have non-comparative knowledge of the two things compared. We compare them in order to see not what each is like but in what respects they are like each other.

What reply should the fallibilist make to this argument? First, what is the content of such an infallible belief? It cannot really be that the way things are looking to me now is pink, since I could be wrong about whether that way is pink or not. What is more plausibly infallible is a belief that things *look that way* now to me. But what does such a belief amount to? What content does it have? It isn't an internal and somehow non-verbal description of how things look; it doesn't say which way they look. All such a belief amounts to is a gesture towards something; and a very strange sort of gesture too, for gestures are normally only comprehensible as public acts with publicly observable objects, while here the gesture is a private act with a private object. If gestures draw one's attention to something, can we see ourselves as somehow drawing our own attention to how things look to us?

The infallibilist may insist that I cannot be wrong in believing that the way things look to me is pink, though I can be wrong in supposing that 'pink' is the word to use to describe the way things look. This is the 'merely verbal error' move. But it seems to misuse the notion of an error which is *merely* verbal. There are several sorts of such errors (a study of the life of Warden Spooner is instructive here, though one can be pardoned the suspicion that some of his errors were deliberate). But the case where, choosing my words carefully with full consciousness of what I am doing, I deliberately pronounce on the nature of my present sensory state is not one of

them. Here if I am wrong, my error is substantial, for in being wrong about whether 'pink' is the word to describe my present experience I am wrong about what pink is and hence about whether my experience is of pink rather than of orange. Here then the error is both verbal and substantial.

If the content of a putatively infallible belief is merely that things are looking that way to me now, there is clearly less room for error than if I were to risk the belief that that way is pink. The less the content, the less the risk, and greater the chance of infallibility. It seems probable, then, that a belief can only be genuinely infallible if it has no content at all. This is the strong fallibilist conclusion. But even if this conclusion is not justified, we can say that infallible beliefs must have vanishingly small content. And the point of this is that the infallible beliefs are intended, within the programme of classical foundationalism, to act as those by appeal to which all others are to be justified. They are the *basic* beliefs which ground all others, our epistemological foundations. And to perform this role they need to have sufficient content to be used rather as premises in inferences. With the reduction in content required to keep them infallible, it seems unlikely that any interesting beliefs about the past, the future, the unobserved or even the present material surroundings could ever be justified by appeal to the basic. Our basic beliefs must have sufficient content to support the superstructure in which we are really interested, and no belief with that amount of content is going to be infallible.

This diagnosis of the infallibilist's errors can be confirmed by considering briefly the arguments of Chisholm, a leading contemporary foundationalist. Chisholm distinguishes between comparative and non-comparative uses of the phrase 'appears white' (Chisholm, 1977, pp. 30 − 3). In the comparative use, 'X appears white' is short for 'X appears the way in which white things normally appear'. But in the non-comparative use, which is found in the sentence 'white things normally appear white', things are different. The latter sentence would be tautologous if we expanded 'appear white' in it as for the comparative use. But it is not tautologous, and hence there must be another non-comparative use of the phrase 'appears white' − a use in which we make a genuine attempt to describe, without comparison, the way in which white things generally appear. And Chisholm claims that in the non-comparative use, appear-statements express what is 'directly evident'. A directly evident proposition is one, in Chisholm's terminology, which is either identical with or entailed by a true contingent proposition which is all

but certain. (A contingent proposition is one which might or might not be true, and might or might not be false.) A belief in a directly evident proposition is not quite the same as an infallible belief, but they share the characteristic with which we are here concerned, that all of them are true.

Chisholm considers various objections to his thesis that there is a non-comparative use of 'appears white' which expresses what is directly evident (and therefore true). Some of his remarks have been echoed in the arguments for infallibilism given above. The last objection he considers runs as follows (Chisholm, 1977, p. 33):

(a) in saying 'Something appears white' you are making certain assumptions about language; you are assuming, for example, that the word 'white', or the phrase 'appears white', is being used in the way in which you have used it on other occasions, or in the way in which other people have used it. Therefore (b) when you say 'this appears white', you are saying something not only about your present experience, but also about all of these other occasions. But (c) what you are saying about these other occasions is not directly evident. And therefore (d) 'this is [*sic* − should be 'appears': JD] white' does not express what is directly evident.

Chisholm comments correctly that in this argument the error is the step from (a) to (b). I agree that this step is invalid. (b) should be (b¹): "When you say 'This appears white', what you say cannot be true unless certain propositions about experiences other than your present ones are true." This creates a much stronger argument, but what interests me here is Chisholm's reason for holding that (b) does not follow from (a). He says:

We must distinguish the belief that a speaker has about the words he is using from the belief that he is using those words to express. A Frenchman [may believe that] "potatoes" is English for apples...; from the fact that he has a mistaken belief about "potatoes" and "apples" it does not follow that he has a mistaken belief about potatoes and apples.

It is clear here that Chisholm is following the traditional line that all apparent error in beliefs about our sensory states is verbal error, and that verbal error is to be distinguished from substantial error. This confirms our original diagnosis of the infallibilist's mistaken move. Chisholm does make that move, and provides no new reason for believing in infallibility.

4.3 FOUNDATIONALISM WITHOUT INFALLIBILITY

In the absence of infallibility, the programme of classical founda-
tionalism collapses. But we saw no reasons to suppose this to be
the only, nor necessarily the best form of foundationalism. What
weaker versions are possible?

The first characteristic foundationalist thesis embodies the
response to the regress argument:

F¹: There are two forms of justification, inferential and non-
     inferential.

But there is a further thesis involved in acceptance of the regress
argument:

F²: Basic beliefs are never justified, even in part, by appeal to
     non-basic beliefs.

F² rules out the suggestion that the non-inferential justification of
basic beliefs is only partial, and needs supplementing by appeal to
other beliefs. This suggestion might tempt those who notice that
although we do normally accept without query a person's descrip-
tion of his own sensory states, we sometimes object by saying, for
instance, 'Surely that traffic light doesn't look orange to you. It's
the top light, and the top lights are always red'. If non-basic beliefs
such as these can reduce the justification of basic beliefs, they can
presumably also increase it, and in this case there would be the
possibility that although our beliefs about our sensory states are
always partially justified just because of their subject matter, they
are never completely or satisfactorily justified unless there is confirm-
ing or at the least a lack of disconfirming evidence at the non-basic
level. But this idea is not available to anybody convinced by the
regress argument. Foundationalists suppose that there are two sorts
of justification, and that the inferentially justified beliefs are justified
by appeal to the non-inferentially justified ones. If we go on to admit
that the latter are partially justified by appeal to the former, we
reintroduce the circle of conditional justification whose sceptical
consequence was that nothing was ever actually justified. Founda-
·tionalists must keep the direction of justification all one-way, from
the non-basic to the basic, or else renounce any use of the regress
argument.

So the only sort of foundationalist who could retain F^1 without F^2 would be one who accepted F^1 for other reasons than those provided by the regress argument. We might hold that our beliefs about our sensory states are always justified to some degree just because of their subject matter (non-inferentially, therefore), whereas most other beliefs are justified inferentially if at all; one could suppose this in an attempt to make sense of the empiricist idea that our beliefs about our present experience have a stability which other beliefs lack, in virtue of which they are able to justify those other beliefs and thus meet the empiricist demand (vaguely expressed here) that all our knowledge be grounded in our experience. This new sort of foundationalist, who will be considered further in chapter 6, could escape the demands of F^2, but only at the cost of abandoning the foundationalists' favourite weapon, the regress argument.

Any foundationalist has a duty to make sense of the possibility that there be non-inferentially justified beliefs. What might these be like? The formal requirements of the regress argument would be satisfied if there were beliefs of any of the three following types:

1   beliefs which are justified by something other than beliefs;
2   beliefs which justify themselves;
3   beliefs which need no justification.

It would be harsh to rule 3 out just because the argument demands that only justified beliefs can justify others; if we can give a reasonable sense to the notion of a belief standing on its own feet, as it were, rather than on those of others, we could easily see 3 as a special case of 2.

We should think of 1 – 3 as formal properties; beliefs with these properties would stop the regress, but any such formal property needs to be grounded in a more substantial 'epistemic' property. Infallibility is such a property, as was remarked earlier. The classical foundationalist supposed perhaps that infallible beliefs would be of type 2. In the absence of infallibility, what similar moves are available?

C. I. Lewis used to claim that basic beliefs were 'certain' or 'incorrigible' (Lewis, 1952 and 1946, ch. 7); it is not always clear whether he thought of these as the same as being infallible or not. In similar vein others (Descartes, perhaps) have held that they are or could be 'indubitable'. We could define incorrigibility and indubitability thus:

A belief is *incorrigible* if and only if no one could ever be in a position to correct it.

A belief is *indubitable* if and only if no one could ever have a reason to doubt it.

Is either of these two properties able to provide us with a slightly weaker but still attractive form of classical foundationalism? In my view, once we have admitted that our beliefs about our sensory states are not infallible and may be false, incorrigibility would be a vice rather than a virtue. The thought of some basic beliefs being incorrigibly false is too horrific to countenance. Equally, we could ask how a fallible belief could be indubitable. For the basic beliefs stand in inferential relationship with more interesting beliefs about public objects, and these beliefs are certainly dubitable; there is always a possibility that someone should have a reason to doubt them. But, if so, it is hard to see how the dubitability of the non-basic beliefs which they support would not rub off onto the basic ones which support them; surely falsehood in a non-basic belief would be a reason to doubt the basic beliefs which support it, once we have admitted that basic beliefs *can* be false.

So neither incorrigibility nor indubitability can provide us with an alternative form of foundationalism. But other possible forms remain. We could suggest that there could be beliefs of type 1 if there were beliefs which were justified by appeal to the facts and that a belief could be so justified if it was caused by the facts. Austin mooted the idea that on occasion what justifies my belief that there is a pig before me is just the pig (Austin, 1962, pp. 115 – 16). But this idea is not easily generalizable unless we suppose that the justification is achieved less by appeal to the pig than by appeal to the fact that there is a pig before me; the fact, in this instance, causes the belief. This move should call to mind some remarks made in 2.4 about the causal theory of knowledge. We accepted there the idea that facts can be causes; the difficulty there that universal facts cannot cause universal beliefs doesn't matter here, because universal beliefs are unlikely to be basic.

Another possible version of foundationalism holds that there are some beliefs which are given us as 'data', and which are fully justified unless something arises to defeat their justification (cf. the use of defeasibility in 2.3). We could call this a 'defeasible' or 'prima facie' justification; it is weaker than that provided by indubitability, because it countenances the possibility that there be reasons against

a basic belief. But it still accepts both F^1 and F^2.

A final version has already been mooted. Weaker than the last, it holds that beliefs given us as 'data' are never fully justified merely for that reason, but that all such beliefs are already partially justified, quite apart from any further support they may receive from other beliefs. Without that further support, however, their justification is insufficient. This is the foundationalism which accepts F^1 but not F^2.

These different versions of foundationalism remain unharmed by the absence of infallibility. In the next chapter we turn to consider a different problem for foundationalism, which may have more damaging effects.

FURTHER READING

C. I. Lewis (1952) and Ayer (1950) defend forms of infallibilism.

Alston (1976) gives a good account of the regress argument, as does Armstrong (1973, ch. 11).

Firth (1964) maps weaker forms of foundationalism from the point of view of its rival, coherentism.

The notion of infallibility used in this chapter needs careful attention. Alston (1971) analyses the differences between infallibility, incorrigibility, indubitability etc.

C. I. Lewis (1952) is a response to Reichenbach (1952) and Goodman (1952), both of whom argue against his view that unless something is certain nothing is even probable.

Most of the above, with many other important papers on topics discussed in this chapter, are reprinted in Chisholm and Swartz (1973).

The foundationalism of Chisholm (1977) explicitly relies on the regress argument at pp. 16 – 20.

The attack on Ayer in Austin (1962) is still well worth reading, especially ch. 10.

Sellars (1963, ch. 5) is an extremely influential attack on the 'myth of the given' embodied in classical foundationalism. This paper is difficult but rewards study.