Realism and the Epistemic Accessibility of Correspondence Truth

Giorgio Volpe

Ph.D. Thesis

The London School of Economics and Political Science

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

A long-standing objection to the correspondence theory of truth is that it is bound to make truth epistemically inaccessible and knowledge impossible. This sort of objection has led many philosophers to espouse anti-realism by subscribing to some kind of epistemic theory of truth. The aim of this thesis is to reject the standard objection against correspondence truth by arguing (i) that no reasonable version of the epistemic theory of truth is going to make truth epistemically more accessible than correspondence truth, and (ii) that in the framework of a naturalistic epistemology correspondence truth can prove sufficiently accessible to our cognitive efforts. Chapter 1 spells out the content of various claims which are usually described as 'realist' and investigates their connections with correspondence and epistemic truth. Chapter 2 introduces the 'inaccessibility' argument against correspondence truth, discusses Hilary Putnam's 'Brains in a vat' purported refutation of 'external' realism, and argues that *ceteris paribus*, every epistemic theory of truth falling short of strict verificationism will fail to make truth epistemically more accessible than a correspondence theory can. Chapter 3 provides a discussion of epistemological internalism. It gives an account of the appeal of epistemological internalism on philosophers in the Cartesian tradition and describes two major theoretical problems it has to face. Chapter 4 focuses on externalist accounts of knowing which make room (or can be modified so as to make room) for the possibility that human beings have, at least in certain circumstances, knowledge of their knowledge. Robert Nozick's 'tracking' analysis of factual knowledge and Fred Dretske's 'information-theoretic' analysis of (perceptual) knowledge are extensively discussed.

Chapter 5 addresses the charge that purely externalist (i.e., naturalistic) accounts of knowing ought to be seen, in Laurence Bonjour's phrase, 'as simply abandoning the traditional idea of epistemic justification or rationality and along with it anything resembling the traditional conception of knowledge'. This leads to a wider discussion of the role and character of epistemic justification in our argumentative practices. Chapter 6 contains a discussion of various sorts of 'naturalized' epistemologies and identifies the 'naturalistic' claims one must be prepared to subscribe to in order to support the thesis that correspondence truth is something human beings can rationally pursue. Finally, a model-theoretic approach to the analysis of the comparative concept of verisimilitude is presented in the Appendix.

Contents

Introduction

The correspondence theory of truth endorsed by such traditional empiricists as Locke and Hume held truth to be a relationship between ideas in the mind and a noumenal world to which we can have no epistemic access. This view of truth burdened philosophers with the huge epistemological task of bridging the gap between our mental representations and the world beyond them. Given the way the problem had been set up, this proved a desperate enterprise, and the theory of truth which had engendered it came to be viewed as increasingly problematic. Today many philosophers believe that the correspondence theory of truth is irredeemably flawed, because they take it to make truth inaccessible and knowledge impossible. And this belief leads them to subscribe to various sorts of disquotational or epistemic theories of truth.

The aim of this work is to argue that the epistemological objections that are often believed to refute the correspondence theory of truth are far from compelling, because the correspondence theory of truth need not engage philosophers in the sort of epistemological exercise that Locke and Hume — and most of the theory's current opponents — take it to involve. It is simply not the case that we can have no epistemic access to the world as it is 'in itself', as opposed to the world as it is 'for us'. There is no inaccessible *noumenon* to which our mental representations ought self-contradictorily to be compared. It is not as if epistemic access to the 'real' world required one to place oneself, as it were, in a God's eye point of view, in order to compare one's own mental representations with reality as it is 'in itself'. I shall argue (i) that to have *access* to reality as it is 'in itself' is, simply, to have correspondence-true representations of it, and

(ii) that to have *epistemic* access to reality as it is 'in itself' is to have representations of it which do not merely happen to be correspondence-true, but which are correspondence-true as the result of an ability to discriminate truth from falsehood within a range of relevant alternatives.

My argument in defence of the correspondence theory of truth will centre on the rejection of epistemological objections concerning its *accessibility* to the cognitive efforts of human beings. I will touch on the semantic and metaphysical issues raised by the *conceivability* of a correspondence relation between linguistic descriptions and a knowledge-independent reality only in so far as this will be required by the necessity of providing a minimal characterization of the relation whose epistemic accessibility I propose to defend. My approach to the issue of the *nature* of this correspondence between linguistic descriptions and reality will be, in other terms, instrumental in the development of my argument for its epistemic accessibility, but I will *not* offer a comprehensive treatment of the semantic and metaphysical issues connected with the conceivability of a correspondence relation between linguistic descriptions and reality.

There is however another respect in which semantic and metaphysical issues are connected with the correspondence theory of truth. This is because commitment to the correspondence theory of truth is often equated to commitment to 'realist' outlooks in semantics and/or metaphysics. I must say that I do not share the belief that the only genuine issue underlying traditional debates about realism is that of the proper interpretation to be given to the statements in a given class. But it is a fact that one's pet theory of what it is for a sentence[1] to be true or false affects what sort of 'things' one will be prepared to grant human beings epistemic access to. The general concern of this work is with that line of reasoning that, starting from the alleged inaccessibility of reality as it is 'in itself', goes on to argue the case for an epistemic theory of truth, and eventually issues in some form of anti-realism. The realist claim that the world we can have epistemic access to is not merely 'phenomenal', but truly knowledge-independent, appears to be jeopardized by the belief that the correspondence theory is bound to make truth inaccessible and knowledge impossible. Accordingly, by providing an argument in

[1] Even if this work is about the epistemological consequences of competing theories of truth, it is totally uncommital on the issue of the 'bearers' of truth. Although terms like 'sentence', 'statement', 'judgement' and 'proposition' are not used as if they were completely interchangeable, no particular significance should be attached to any of their occurences. For example, the fact that I describe theories of truth as theories about what it is for a *sentence* to be true or false should not be taken to mean that I endorse the view that it is *sentences* that are the bearers of truth.

defence of the correspondence theory of truth I hope to undermine what is usually taken to be one of the most compelling reasons for believing that full-blown realism is no longer a viable metaphysical option (needless to say, this does not mean that I will establish the *truth* of full-blown realism).

Showing that the correspondence theory of truth need not make truth inaccessible and knowledge impossible is not, of course, the only way of defending the realist claim that the world to which we can have epistemic access is a truly knowledge-independent world. A number of philosophers have argued that one can reject the view that truth is correspondence to a knowledge-independent reality without thereby being committed to subscribe to an epistemic theory of truth, let alone to anti-realism. The way in which this is alleged to be possible is by endorsing a disquotational (or redundancy) theory of truth, the content of which is supposedly neutral with respect to the issue of realism. My opinion about the disquotational theory of truth is that it cannot be a viable alternative to the correspondence theory because it cannot make sense of the goal-directedness of our cognitive efforts. I will sketch an argument to this effect in section 1.4, but an extensive discussion of the disquotational theory of truth and of its relationship to realism is beyond the scope of this work. Accordingly, my defence of realism *via* the correspondence theory of truth will have to be seen as exploring just *one* possible strategy for the defence of realism (i.e., the strategy dictated by the assumption that the disquotational theory of truth is unworkable).

What follows can be described as an attempt to spell out the claims that someone following this strategy must be prepared to subscribe to in order to develop a conception of 'epistemic access' consistent with the belief that human beings can and do know many things about the world as it is, and not merely about the world as it appears to them. Chapter 1 contains an analysis of various forms of realism and a discussion of their relationship with the correspondence and epistemic theories of truth. Chapter 2 introduces the 'inaccessibility' argument against the correspondence theory of truth and argues that epistemic theories of truth are affected by the same epistemological problems that are supposed to haunt the correspondence theory. Chapter 3 offers a characterization of internalist and externalist positions in the theories of knowledge and epistemic justification and exposes some difficulties in the traditional approach to epistemology. Chapter 4 discusses the analyses of knowledge of two leading externalist epistemologists, Robert Nozick and Fred Dretske. Dretske's information-theoretic approach to knowledge

is expanded and tentatively endorsed. Chapter 5 attempts to reject the most common objections to an externalist account of knowledge by developing a pragmatic view of (internal) epistemic justification as an activity whose purpose is to affect other people's as well as one's own beliefs. Chapter 6 contrasts the traditional project of a priori epistemology with a naturalistic approach to the vindication of cognitive methods and argues that only the latter is a viable and promising enterprise.

The project of defending realism by showing that the correspondence theory of truth need not make truth inaccessible and knowledge impossible leads me to discuss 'Brains-in-a-vat' thought-experiments as well as the possibility of attaining knowledge of one's knowledge. This might create the impression that my aim is to address the central issues of traditional epistemology, which is not. Even though part of what I have to say does have some bearing on those issues, my aim is not to provide a solution for the major questions of epistemology as they have been traditionally understood. I promise no refutation of (radical) scepticism, because I believe that the *epistemological* attempt to bridge the gap between our mental representations and the world beyond them by means of a priori considerations is a desperate and useless enterprise. The thrust of this work is that the *semantic* attempt to bridge the gap between our mental representations and the world beyond them by reinterpreting our conception of truth in an anti-realist fashion is equally pointless.

Realism and Truth

The aim of this chapter is to spell out the content of various claims which are commonly described as 'realist' and to investigate their connections with the two theories of truth which are standardly associated with the realism/anti-realism debate, i.e., the theory of truth as correspondence to a knowledge-independent reality and various sorts of epistemic theories of truth.

Sections 1.1 and 1.2 raise doubts about the popular view that the point of most realism/anti-realism debates can and should be captured in purely semantic (Dummett), or perhaps epistemological (Papineau), terms. Section 1.3 describes the metaphysical thesis which I think underlies realist claims both of an epistemological and a semantic sort. Section 1.4 provides a minimal characterization of the correspondence theory of truth and discusses its relationship with Tarski's semantic definition of truth. Section 1.5 introduces epistemic truth and explains why a successful defence of correspondence truth is important for realism even if the content of realism cannot be straightforwardly identified with commitment to a correspondence theory of truth.

1.1 Is Realism a Doctrine about Meaning?

Current realist doctrines can be seen as falling into two very general kinds. On the one hand, there are those realist doctrines which make one or more of the following claims, namely, (i) that the world *would* exist and retain its structural properties even if it were

the case that there were no cognizers around, (ii) that the world *could* be utterly different to the way we take it to be on the basis of our best (scientific) theories, and (iii) that certain statements about the world *could* be true even if no human being should ever be in a position to verify them. On the other hand, there are those realist doctrines which claim that the world *is*, more or less, the way we take it to be on the basis of our best (scientific) theories.

Doctrines falling under one kind are clearly not incompatible with doctrines falling under the other, but this does not prevent them from saying quite different things. So one might wish to classify the realist doctrines according to their modal status and say that there are *realisms-with-modality* and *realisms-without-modality*. Realism-without-modality is what currently goes under the label of *scientific* realism, whereas realism-with-modality covers various kinds of *metaphysical*, *epistemological*, and *semantic* realisms.

Scientific realism (realism-without-modality) has been occupying centre stage in the philosophy of science for about twenty years. Following Richard Boyd, Putnam (1976, 179) describes scientific realism as affirming (i) that the laws and theories currently accepted in mature science are at least approximately true and (ii) that theoretical terms occurring in those laws and theories successfully refer to entities in the world. Scientific *anti*-realism, on the other hand, denounces both these claims as seriously undersupported by scientific evidence. Van Fraassen (1980, ch. 1), for example, contends that all that one can legitimately believe about the laws and theories of mature science is that they are 'empirically adequate', i.e., that they *save the phenomena* they are purposed to describe.

Two more claims are frequently taken to be part of the content of scientific realism, namely, the convergence thesis and (some minimal version of) the correspondence theory of truth. The convergence thesis holds that 'the historically generated sequence of theories of a mature science is a sequence of theories which are improving in regard to how approximately true they are' (Newton-Smith 1981, 39). A minimal version of the correspondence theory of truth holds at least that 'to be true (false) is to be true (false) in virtue of how the world is independently of ourselves' (Newton-Smith 1981, 29; the promise of a more detailed characterization of the correspondence theory of truth will be fulfilled in section 1.4). However, while the convergence thesis is in fact shared by most self-avowed scientific realists, the

correspondence theory of truth is not. Among the scientific realists who reject the correspondence theory of truth are Hilary Putnam, who now believes it to be 'incoherent' (Putnam 1983, 230), and Brian Ellis, who has argued that scientific realism is incompatible with every form of the correspondence theory of truth because the ontology of science cannot possibly accommodate the entities to which that theory assigns the role of the *bearers* of truth (see Ellis 1990, ch. 5). On the other hand, Bas van Fraassen, whose 'constructive empiricism' provides one of the most compelling forms of (scientific) anti-realism currently available, appears non-equivocally committed to the correspondence theory of truth (van Fraassen 1980, 90, 197), the acceptance of which can be argued to be instrumental to the success of his argumentative strategy. So I think it wiser to make our characterization of scientific realism logically independent of the correspondence theory of truth. Further reasons for this move will emerge later in the discussion.

The sort of realism with which I will be centrally concerned in this work is not scientific realism, but realism-with-modality. As the opening paragraph of this section suggests, it is possible to distinguish three different though closely related theses, various combinations of which can be aptly described as realism-with-modality. First of all, there is the old-fashioned *metaphysical* thesis about the existence of a knowledge-independent reality. Secondly, there is the *epistemological* thesis that our judgements might always fail to represent the world as it is. Finally, there is the *semantic* thesis that the statements in certain classes could be true even if (in some sense to be better specified) unverifiable[1]. In this section and the next one I shall argue that none of these theses is parasitic upon, nor reducible to, any of the others.

The most influential recent work on realism-with-modality is to be found in the writings of Michael Dummett. The semantic thesis that the statements in certain classes could be true though unverifiable is of course very close to Dummett's own definition of realism. He contends in fact to have replaced the old metaphysical issue of realism with a much more interesting semantic issue about the conception of truth required to make sense of our linguistic practices.

---

[1] Horwich (1982) has a similar distinction between 'three forms of realism', which is, however, far from equivalent to the one I propose here. His 'epistemological' realism is in fact my realism-without-modality; his 'semantic' realism combines features of my semantic and metaphysical realism; and his 'metaphysical realism', which is a theory about truth's being a primitive non-epistemic idea, entails but is not entailed by my metaphysical realism.

Dummett claims to have discovered the common features shared by many different debates commonly regarded as involving the issue of realism (Dummett 1978, 45). In his view, the realism-nominalism controversy in the theory of universals, the realism-idealism controversy in metaphysics, and many other philosophical controversies, like those about the reality of the past and the existence of mathematical entities, all are essentially *semantic* in character. Any of these controversies concerns the interpretation to be given to the statements in some given class. In each case, the hallmark of realism is endorsement of the thesis that statements in the relevant class are determinately either true or false independently of our being capable of recognizing their truth-value.

Dummett himself does not stress the modal character of his semantic statement of realism, because he focuses on statements (e.g. about the distant past) that are in fact undecidable, thus implying that there is no real disagreement between the realist and the anti-realist as far as decidable statements (e.g. about the primary qualities of observable physical objects) are concerned. It has been argued, though, that if Dummett's general argument against realism is that we may have no *bona fide* understanding of statements having unverifiable truth conditions, he must then deny not only that unverifiable statements are (objectively) true or false, but also that currently decidable statements *could* be true even if they were not, as in fact they are, verifiable by us (see Loar 1987, 84-86). Dummett has not objected to this way of formulating his position, and though nothing really substantive hinges on this issue, I shall ascribe to him the thesis that realism can be described as the modal claim that 'for all *s* of such and such a class, it could happen that *s* were true although *s* were unverifiable' (Loar 1987, 83). This seems to me the sharpest way to spell out the realist's commitment to what is usually called an 'evidence-transcendent' notion of truth. Dummett's anti-realism can be aptly described as the denial that genuinely meaningful statements could be true though unverifiable.

Running together in a rather cavalier way metaphysical realism and the correspondence theory of truth, Dummett points out that the commitment to an evidence-transcendent notion of truth apparently 'involves' the existence of a knowledge-independent reality in virtue of which the statements in some given class are either true or false (Dummett 1982, 55). This appearance is supposed to explain why the doctrine of realism has traditionally been couched in metaphysical terms, as the thesis that the relevant portion of reality would exist and retain its structural properties even if it were the case that there were no cognizers around. But talk about the existence of a

knowledge-independent reality, Dummett suggests, is merely metaphorical (Dummett 1978, xxv-xxvi, 229) and can be usefully dismissed in favour of the semantic thesis from which it derives all its content. In short, Dummett's contention is that realism and anti-realism are primarily theses about *meaning*, and only derivatively (and metaphorically) theses about *what there is* (Dummett 1978, xl). Dummett claims a greater generality for his semantic approach, on the ground that it encompasses controversies (e.g. about the future and about ethics) which can hardly be construed as metaphysical controversies about the existence of entities of a given kind (Dummett 1982, 55).

Although the general argument of this work shall not depend crucially on the rejection of Dummett's approach, I shall now briefly explain why I am not convinced by his claim that the epistemological, and especially the metaphysical, versions of realism are parasitic upon the semantic claim that the statements in certain classes could be true though unverifiable.

Dummett's only argument for his distinctively semantic approach to the realism/anti-realism debate is the allegedly metaphorical character of metaphysical realism. Dummett himself has often presented his anti-realism as a generalisation of mathematical intuitionism. Thus Michael Devitt is likely to be right when he points out that Dummett's thesis on the metaphorical character of metaphysical talk about reality arises from his philosophy of mathematics (Devitt 1991b, 264-266). In that particular field, it may be sensible to suppose that platonic talk about mathematical objects is purely metaphorical, as the real problem appears to be, in Kreisel's words, 'not the existence of mathematical objects, but the objectivity of mathematical statements' (quoted in Dummett 1978, xxviii). However, there is no reason to believe that metaphysical talk about the existence, for example, of physical bodies should inherit the supposed metaphorical character of platonic talk about mathematical entities. After all physical bodies, unlike platonic entities, are supposedly capable of bringing about causal effects on human beings. This is why we can gain epistemic access to physical facts. We can have evidence of the existence of physical bodies because they are causally efficacious, whereas the causal inefficacy of mathematical objects suggests that we can hardly have any reason to believe in their existence.

This asymmetry surely provides a *prima facie* case for resisting a straightforward assimilation of the metaphysical talk about the existence of knowledge-independent physical bodies to the (allegedly) metaphorical talk about the existence of knowledge-

independent platonic entities. But if no other arguments are offered to support the thesis of the purely metaphorical character of metaphysical realism, this need not be seen as deriving its whole content from semantic realism, and the question of the logical relationship between the two theses becomes an open issue.

Now, even if one may be inclined to believe, Dummett-style, that the metaphysical and epistemological theses follow from the claim that statements about the world can be true though unverifiable, no entailment appears to obtain in the opposite direction. If it makes sense to ask whether the world *could* exist and retain its structural properties even if it were the case that there were no cognizers around, then it seems to me that one could well be a metaphysical realist *and* a semantic anti-realist.

The first argument to this effect comes from the realization that a metaphysical realist can legitimately claim that the realm of what there is is (much) larger than the realm of what we can think about or conceive of (this claim is endorsed, for example, by Popper 1972, ch. 5 and Nagel 1986, ch. 6). A metaphysical realist who believed that reality cannot be completely described by means of our concepts could subscribe to the thesis that no genuinely meaningful statement can be true though unverifiable[2] without denying that the world could exist and retain its structural properties even if it were the case that there were no cognizers around. She could do so by contending that although no meaningful statement is genuinely unverifiable, still the true structural properties of the world are such that they cannot even be expressed in our language (or in a conceivable extension of it). This contention would enable her to say that the existence of a structured world does not depend on the availability, let alone the verifiability, of true linguistic descriptions of its nature.

The second argument against the claim that metaphysical realism entails semantic realism is as follows. Take a person who is so confident about her own beliefs as to be prepared to assert, for any sentence *p*, whether *p* or not-*p*. Such a person might endorse, for example, a very credulous (and conservative) epistemology, according to which the mere fact that *p* is believed by S (that is, by herself) is a sufficient condition for *p*'s truth (and a sufficient reason for S's accepting it), as long as *p* does not appear to contradict other propositions previously accepted by S (in which case it ought to be rejected; we may assume S to be logically omniscient). Such a person might even be willing to defend the reliability of her epistemological stance by claiming that she is never mistaken

---

[2] This of course is neither Popper's nor Nagel's case.

because, say, God ensures that her beliefs are unfailingly true. Of course such a person would be very unlikely to end up with any substantive body of knowledge about the world, but here the cognitive effectiveness of her epistemology is not at issue. What is relevant is rather that she could accept metaphysical realism without being committed to ascribing evidence-transcendent truth conditions to any statement at all. For I can see no reason why the fact that no statement could be, according to her view, unverifiably true should commit her to deny that the world could exist (and retain its structural properties) even if it were the case that there were no cognizers around. And this goes to show that as soon as metaphysical realism is regarded as being a meaningful claim on its own, it is logically consistent with the denial of semantic realism.

This example draws our attention to the extent to which one's attitude towards the relationship between the metaphysical and the semantic thesis is shaped by one's previous epistemological commitments. Commitment to the objective existence of a knowledge-independent reality underdetermines people's attitudes towards semantic realism and/or anti-realism. By accepting a dogmatic epistemology of the sort described in the last paragraph, one can be at the same time a metaphysical realist and a semantic anti-realist.

These considerations seem to suggest that Dummett's semantic approach cannot be the whole picture of the realism issue. Semantic realism may be a sufficient, but certainly not a necessary condition for metaphysical realism. Therefore, the fact that semantic realism apparently entails metaphysical realism need not be taken as a sign of the former's being any more ultimate than the latter. It is far more sensible to take this fact, as A.C. Grayling has suggested, as a sign of the unavoidable metaphysical underpinning of semantic realism:

> On Dummett's order of exposition, if one accepts a commitment to bivalence and knowledge-independence of truth-value, one is thereby committed to holding that a knowledge-independent reality exists which makes statements determinately true or false. In this way the thesis about truth and its semantic embedding appears to be the decisive factor. But the *logical* order of dependence among these commitments is [...] the reverse of his order of exposition. The crucial commitment is to there being a knowledge-independent reality or realm of entities, for, without this view already in place for the semantic thesis to presuppose it, that thesis is empty.
>
> (Grayling 1992, 52)

Of course some philosophical positions which come out as realist on the basis of Dummett's criterion turn out to be totally independent of any commitment to the existence of a knowledge-independent reality. But from this fact one may wish to conclude that 'what we should say about those "realisms" which are not readily classifiable in terms of some reality or realm of entities is, simply, [...] that they are not *realisms*' (Grayling 1992, 54). The resulting picture will be less unified than that provided by Dummett's purely semantic approach, but unification *come what may* is not necessarily a cognitive virtue.

1.2 Is Realism a Doctrine about Knowledge?

Even if Dummett's unilateral approach may be seen, in Devitt's phrase, as putting the semantic cart before the metaphysical horse, his emphasis on the role played by the theory of truth in the characterization of realism has become commonplace in recent philosophy. This may partly be explained by the feeling that old-fashioned, metaphysical realism about the 'external' world must be either a boring truism or an ontological extravagance. However this may be, it is a fact that in recent Anglo-Saxon philosophy '"Realism" represents, in one way or another, the idea that truth is independent of us' (Haack 1987, 276).

David Papineau, for example, rejects Dummett's semantic characterization of the realism/anti-realism debate, but nevertheless regards this debate as concerned with what sort of truth, if any, can be expected from our theories about the world. Papineau's description of realism-with-modality is unequivocally epistemological:

> realists think there is always a possibility that our judgements might fail to represent the world as it is.
>
> (Papineau 1987, 2)

Papineau's statement is to be construed as saying that not only our ordinary beliefs, but even our best theories about the world could in fact be false. A similar idea is expressed by Crispin Wright, who writes that the metaphysical realist is committed 'to the possibility that even an ideal theory might be false or seriously incomplete' (Wright 1992, 91), and plays a central role in Pettit's (1991, 590; 592) somewhat more

complicated version of realism-with-modality. Davidson (1990, 308) attributes to (one of the many) Hilary Putnam(s) the idea that (metaphysical) realism is the view 'that all our best researched and established beliefs and theories may be false' (Davidson himself is, in this sense, an anti-realist).

Papineau suggests that it is because the realist believes that there is some thought-independent world that she also believes that our best theories of the world could eventually fail to be true:

> Intuitively, realists are philosophers who accept that there is an independent reality which is as it is independently of human judgement. But because they think of reality and judgement as separate in this way, realists think there is always a possibility that our judgements might fail to represent the world as it is.
> (Papineau 1987, 2)

However, the reference to 'an independent reality' remains in the background, playing a merely 'pedagogical' role. Like many other writers after Dummett, Papineau seems to believe that metaphysical talk about the existence of a knowledge-independent reality is either empty or hopelessly vague, and therefore attempts to capture the content of realism in purely epistemological terms. This raises the question, Can a purely epistemological characterization of realism do justice to our pre-theoretical intuitions any more than Dummett's purely semantic characterization? My view is that epistemological realism, like semantic realism, is at most a sufficient, but not a necessary condition for metaphysical realism. In Descartes's *Principles of Philosophy* we find a position that Papineau's epistemological characterization of realism seems unable to handle properly.

> Suppose [...] that we use only principles which we see to be utterly evident, and that all our subsequent deductions follow by mathematical reasoning: if it turns out that the results of such deductions agree accurately with all natural phenomena, we would seem to be doing God an injustice if we suspected that the causal explanations discovered in this way were false. For this would imply that God had endowed us with such an imperfect nature that even the proper use of our powers of reasoning allowed us to go wrong.
> (Descartes 1984-85, I, 255)

Descartes believed in the objective existence of a knowledge-independent reality (Williams 1978, 198 ff. argues that he also had a correspondence theory of truth) but, trusting his benevolent God, he did not believe that the theories about that knowledge-

independent reality which had been the outcome of the most careful application of his methodology could ever turn out (say, in the after-life) to be false. On the other hand, since he openly recommended withdrawing judgment rather than accepting insufficiently supported beliefs, Descartes was clearly committed to regard his error-proof methodology as incapable of deciding the truth-value of a considerable number of statements, thus qualifying as a semantic realist.

If one applies Papineau's definition to Descartes, the French philosopher will come out as an anti-realist who believes that the results of the painstaking application of his God-warranted methodology cannot be mistaken. But can this be the whole story about Descartes' attitude towards realism? Certainly it does not take into account his semantic anti-realism. And the reason why Descartes believes that our best theories about the world cannot fail to represent the world as it is is not that there is no 'independent reality which is as it is independently of human judgement', but God's benevolence. The anti-realist verdict yielded by Papineau's criterion cannot be right if his epistemological definition is not to contradict those very pre-theoretical intuitions he relies on in order to bestow some plausibility on his official characterization of realism.

Now, it is true that dogmatic epistemologies — i.e., epistemologies claiming that human beings have access to (recognizably) infallible cognitive methods and that the certainty guaranteed by such methods is an essential feature of knowledge — have lately become rather unpopular. As a consequence, Papineau's definition will usually provide more appropriate results than with Descartes. But this does not suffice to dispel the feeling that, if one had not been so indelibly impressed by Dummett's dismissal of the metaphysical thesis as merely metaphorical, one should find it reasonable to take a further step and openly address the metaphysical aspect of realism which seems to underlie both the epistemological and the semantic theses.

1.3 Realism as a Doctrine about the Existence of a Knowledge-Independent Reality

The upshot of the previous sections is that both semantic and epistemological anti-realism can be taken to be reliable indicators of a more fundamental anti-realist attitude only if certain *epistemological* questions have been previously settled in an admittedly sensible,

but by no means mandatory, manner. This is why the old-fashioned *metaphysical* question of the existence of a knowledge-independent reality cannot be so easily dismissed. In other words, unless one is prepared to reject several kinds of dogmatic epistemologies, describing a person as a semantic or epistemological anti-realist will leave unsettled her *metaphysical* position.

Metaphysical realism should be regarded, I think, as an independent and irreducible claim, apparently boring and trivial, but in fact often denied, which lies at the foundation of every other claim in the realism-with-modality family. Metaphysical realism is not the same as realism about the 'external' (natural) world. As Devitt (1991, 15) points out, one could be a metaphysical realist about mental states without being a metaphysical realist about the natural world. For a person who rejects, as several philosophers are today inclined to do, both the 'incorrigibility thesis' and the 'self-intimation thesis' about the mental[3], the question whether a mental event exists will be a question about something altogether objective.

Old-fashioned realism about the external world is a metaphysical doctrine about the objective (i.e., knowledge-independent) existence of non-mental reality. The objectivity requirement is evidence of an epistemological element within metaphysical realism: those entities which are asserted to exist must exist independently of any actual or possible knowledge of them (which does not mean that they can never become the objects of any sort of knowledge). The modal character of metaphysical realism in general, and of metaphysical realism about the external world in particular, is apparent as soon as the objectivity requirement is spelled out as the subjunctive claim that even if it were the case that there were no cognizers around, reality would not *ipso facto* cease to exist or loose its structural properties. (Granted, since most cognizers are also agents, if it were the case that there were no cognizers around, some aspects of reality would certainly be different. But it would be the absence of those cognizers' knowledge as knowledge-of-agents, rather than as knowledge-of-cognizers, that would make the difference. The kind of knowledge-dependence the metaphysical realist about the external world wants to deny has nothing to do with those trivial causal relations by which such mental states as beliefs and desires can affect — through the intervention of human action — the physical world. What she wants to deny is rather that kind of

[3] The 'incorrigibility thesis' is that a person cannot be wrong about, and the 'self-intimation thesis' that she cannot be ignorant of, her own mental states.

epistemological knowledge-dependence where the knowing subject *constitutes* the object of her knowledge and can have epistemic access only to such a *phenomenal* object as opposed to the object as it is in itself).

Metaphysical realism about the natural world is of course one of the most popular forms of metaphysical realism (and one which I definitely want to subscribe to), but the arguments developed in this work are not particularly concerned with the existence of 'external' reality. My discussion of realism and correspondence truth will have a bearing on metaphysical realism about the natural world only in so far as it will concern the issue of metaphysical realism in its utmost generality.

1.4 Correspondence Truth and Tarski's Semantic Definition of Truth

I have argued that commitment to an evidence-transcendent notion of truth is at most a sufficient, but not a necessary condition for (metaphysical) realism. One can be a (metaphysical) realist without subscribing to an evidence-transcendent notion of truth (or, for that matter, without taking the view that even our best theories about the world could in fact be false). If this approach is correct, then that paradigm of evidence-transcendent notion of truth which everybody has in mind when discussing these issues — truth as correspondence to a knowledge-independent reality — cannot itself be properly described as evidence-transcendent independently of the particular epistemological framework in which it is put to work. Within the dogmatic epistemological framework described in section 1.1, for example, correspondence truth proves anything but evidence-transcendent. On the other hand, within a sceptical framework correspondence truth provides the most formidable weapon for proving that truth lies beyond human reach. In itself, the correspondence theory of truth is a theory about what makes our sentences either true or false, but is absolutely silent about any question relating evidence and truth.

What does the correspondence theory of truth say? Unlike in the days of Wittgenstein's and Russell's logical atomism, in its most updated versions the correspondence theory of truth does not say that true sentences 'picture' or are 'isomorphic' to knowledge-independent facts or states of affairs. This is the minimal

characterization of correspondence truth that I shall use throughout this work, which is a modified version of Devitt's (1991b, 29) characterization:

> Correspondence Truth (x): Sentences of type x are true or false in virtue of: (1) their structure; (2) the referential relations between their parts and reality; (3) the knowledge-independent nature of that reality.

I have kept Devitt's relativization to sentence-types because I do not want to give the impression that I am supporting the view that sentences about moral or aesthetic matters are either capable of being *correspondence*-true or false, or neither true nor false.

The relationship between the correspondence theory of truth and Tarski's semantic theory of truth has long been debated. Several contemporary champions of correspondence truth believe that Tarski's semantic definition has succeeded in rehabilitating the traditional view of truth as correspondence (see Popper 1963, 223 f. and Zahar 1984, 165; they draw support for their view from Tarski 1944, 342 f.). This is controversial (see, e.g., Haack 1976), and Tarski's own remark that 'we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians — whatever we were before' (Tarski 1944, 362) is frequently taken to undermine this construal of his achievement. Many writers have recently embraced the view that Tarski's theory is either a sophisticated version of the disquotational theory of truth or philosophically uninteresting (see Davidson 1990, 288). I shall say only a few words about these slippery issues, which are not essential to the main argument of this work. First of all, I shall briefly describe the content of the disquotational theory of truth. Then I shall explain why I am inclined to believe that it cannot constitute a viable alternative to the correspondence theory. Finally, I shall try to assess to what extent Tarski's semantic definition can be seen as providing something more than a merely disquotational theory of truth.

The first formulation of the disquotational (or redundancy) theory of truth is usually credited to Frank Ramsey, but the idea can be found already in Frege's work; a thorough defence of the theory is developed in Horwich (1990), who calls it the 'minimalist' theory. The gist of this theory is that saying that a sentence is true is just another way of asserting it or, in other terms, that everything one can say about the concept of truth is expressed by asserting all the instances of the 'disquotational' schema,

'*p*' is true iff *p*

(this claim is sometimes called the 'equivalence thesis'). According to the disquotational theory of truth, saying that *p* is true does not involve attributing the property of being true to *p*; the predicate 'true' is, in Quine's phrase, just an instrument of 'semantic ascent', i.e., a device which enables us to make meta-linguistic claims about the sentences of our language.

The reason why I am inclined to reject the disquotational theory of truth is, as I said in the introduction, that it cannot make sense of the goal-directedness of our cognitive efforts. For it deprives of content the assertion that their aim is, roughly, to ensure that we accept that *p* if and only if it is true that *p* (this characterization of the aims of our cognitive efforts is of course a gross oversimplification, but it will do for the moment). My claim is that, by depriving of content the assertion that the aim of our cognitive efforts is to ensure that we accept that *p* if and only if it is true that *p*, the disquotational theory of truth becomes inapplicable. The basis for this claim comes from Blackburn's (1984, 229-233) rejection of Frege's 'blockbuster' argument against the possibility of a genuine analysis (i.e., an analysis providing a 'substantive' conception) of truth. The point is that the disquotational theory of truth cannot make sense of the distinction between genuine assertions, which are epistemically responsible in the sense that they involve the speaker's commitment to provide an epistemic justification of their content, and such epistemically irresponsible utterances of *prima facie* descriptive sentences as can be found in dramatic or ludic contexts.

The existence of contexts in which the utterance of *prima facie* descriptive sentences does not involve the speaker's commitment to provide an epistemic justification of their content shows that *saying that p* is not always *saying that p is true*: it is only *asserting that p* that is always equivalent to *asserting that p is true*. But this means that the disquotational theory of truth cannot do the job it is claimed to do — indeed, it cannot even be ascribed a determinate content — unless one already knows how to tell those epistemically irresponsible 'sayings' that provide no counterexample to the equivalence thesis from those genuine assertions that involve the truth of a proper instantiation of the disquotational schema. And the only way in which one can hope to tell epistemically irresponsible 'sayings' from genuine assertions is, as far as I can see, by appealing to their *goals*. If *p* is uttered, say, at the afternoon performance of the

Royal Shakespeare Company, *saying that p* fails to be equivalent to *saying that p is true* because dramatic performances are not directed to ensuring that actors say that *p* if and only if it is true that *p*. But if it is false that an actor's saying that *p* is equivalent to his saying that *p* is true, it will be impossible to explain the fact that *asserting that p* is always equivalent to *asserting that p is true* by claiming that the locution 'it is true that *p*' does not mean anything more than or different from '*p*'. The reason why *asserting that p* is always equivalent to *asserting that p is true* is, rather, that the aim of our cognitive efforts is to ensure that we accept that *p* if and only if it is true that *p* (I am assuming that [honestly] asserting that *p* presupposes *accepting* that *p*). But this can count as an explanation of the fact that *asserting that p* is always equivalent to *asserting that p is true* only if S's saying that it is true that *p* is *not* taken to be equivalent to S's asserting that *p*. What is required to make sense of the distinction between epistemically irresponsible 'sayings' and genuine assertions is then a *substantive*, and not merely a disquotational, theory of truth. But if having a substantive conception of what it is for a sentence to be true or false is necessary to ascribe a determinate content to the equivalence thesis in the first place, it seems clear that this cannot be harnessed to establish that truth is no property of sentences, but merely an instrument of 'semantic ascent'.

I am not fully confident that a properly worked out version of the argument I have just sketched could provide a strict refutation of the disquotational theory of truth, but, as I said in the introduction, an extensive discussion of this theory is far beyond the scope of this work. Moreover, it is surely not necessary for the purpose of this section, which is to discuss the relationship between the correspondence theory of truth and Tarski's semantic definition. To this I shall presently devote my attention.

The now popular point that Tarski's definition cannot provide anything like a substantive theory of truth for a given language because the instances of Tarski's (T) schema (S is true iff *p*) turn out to be mere tautologies is not conclusive. As Davidson (1990, 292-295) has pointed out, the explicit aim of the definition is, in Tarski's own words, 'to catch hold of the actual meaning of an old notion' (Tarski 1944, 341), rather than to stipulate a new meaning for an old world, and his convention-T would have no point if it were not meant to guarantee the conformity of his definition to our intuitive concept of truth. Therefore nothing prevents us from interpreting the instances of

Tarski's (T) schema as empirical truths about a real language rather than as tautologies about an arbitrarily defined ideal language.

Tarski's definition does not specify what *kinds* of entities must exist so that the truth-value of the sentences in our formalised language may be determined. Therefore Tarski's definition is compatible with different ontologies (e.g., with phenomenalism as well as with materialism), and does not require the existence of an external, physical world. But satisfaction is a semantic relation obtaining (or failing to obtain) between sentences and sequences of objects. *If* the objects involved are understood as knowledge-independent, *then* Tarski's definition appears to provide a clear sense in which a sentence in a formalised language can *correspond* (or fail to correspond) to knowledge-independent reality. The question is, Can Tarski's theory accommodate, or even support, the view that the objects in the sequences should be conceived as being knowledge-independent? The fact that Tarski's definition of satisfaction for atomic open sentences is merely *enumerative* (roughly: 'The sequence of length one consisting of just x satisfies the atomic open sentence "y is a tree" if and only if x is a tree') does raise some doubts about this. For one can suspect that Tarski might be relying on a merely disquotational theory of reference (that is, a theory according to which everything one can say about reference are such trivialities as: 'Paris' denotes Paris, 'tree' refer to trees, etc.), which of course would cast serious doubts on the success of his attempt to provide a *correspondence* analysis of truth. A disquotational theory of reference is in fact perfectly compatible with Quine's and Putnam's view that 'Paris' may be said to refer to Paris, and 'tree' to trees, only *within one's own conceptual scheme*. Accordingly, if the enumerative character of Tarski's definition of satisfaction for atomic open sentences was a consequence of his adopting a purely disquotational theory of *reference*, it would be very difficult indeed too regard his definition as providing anything more than a disquotational theory of *truth*.

Furthermore, it can be argued that Tarski's enumerative definition of satisfaction for atomic open sentences fails to provide a truly general definition of satisfaction, as opposed, say, to different definitions for satisfaction-in-L_1, satisfaction-in-L_2, etc. This in turn means that Tarski's definition of truth fails to provide a truly general definition of truth, as opposed to different definitions of truth-in-L_1, truth-in-L_2, etc. And this failure cannot be easily explained away as a consequence of the trivial fact that a sentence's being true or false depends on what language that sentence is a sentence of.

What Tarski's definition fails to do is to say *what the truth predicates of various languages have in common*. Tarski's definition fails to provide an account of our ordinary, universal notion of truth, as opposed to an account of several unrelated notions of truth-in-L_1, truth-in-L_2, etc.

Field (1972) has argued that Tarski's definition fails to provide a physicalistically acceptable definition of truth precisely because it employs a merely enumerative definition of satisfaction, which can only guarantee an extensional equivalence of *definiens* and *definiendum*, as opposed to a true reduction of the former to the latter. Field claims that Tarski's definition must be complemented by a physicalistic theory of reference if it is to give a philosophically unexceptionable explanation of the connection between language and (extralinguistic) reality. The replacement of Tarski's enumerative definition of satisfaction with a physicalistic theory of reference would (i) justify the claim that Tarski's definition can provide a truly general definition of truth and (ii) dispel any suspect that Tarski's definition might be in fact nothing more than a sophisticated version of the disquotational theory of truth.

I do not know whether a physicalistic theory of reference of the kind envisaged by Field has any chance of ever being developed. Moreover, even if Field is right in rejecting 'semanticalism', i.e. the claim that there are irreducibly 'semantic' facts, what is needed to justify the claim that Tarski's machinery can provide a truly general definition of truth is merely a truly general *theory* of reference, which need not necessarily be a physicalistic one (Niiniluoto 1987, 140). The upshot of this brief discussion then seems to be that, given the unavailability of a general (physicalistic or other) theory of reference, Tarski's definition of truth cannot be regarded as providing an adequate analysis of 'the actual meaning of an old notion' because it fails to provide a truly general correspondence *theory* of truth. However, Tarski's definition need not be regarded as purely disquotational either, because — Quine and Putnam notwithstanding — there does not seem to be any conclusive argument for the impossibility of a general (physicalistic or other) theory of reference to knowledge-independent objects, let alone for the impossibility of *reference* to knowledge-independent objects. If such a general theory should become available, Tarski's definition complemented by this theory might very well be regarded as providing a rigorous analysis of truth as correspondence to a knowledge-independent reality. Meanwhile, I think it can safely be assumed that if Tarski's machinery cannot be seen

as fleshing out our (minimal) characterization of correspondence truth, it need not be seen as providing an alternative to it either.

However, even if Tarski's definition is seen as a first step towards a fully satisfactory correspondence theory of truth rather than as a self-contained disquotational theory, there is one more point which should be emphasized. Tarski's theory must not be taken to imply the further claim that the facts described by the sentences on the right-hand-side of any given instance of the (T) schema are the knowledge-independent entities that make the sentences named on the left-hand-side either true or false. Tarski's theory lends no support to the view that a sentence is true iff there is some knowledge-independent fact to which it corresponds. Indeed, what makes a sentence true is not, on Tarski's analysis, a specific fact or state of affairs, but sequences of objects. If one were to accept the argument recently rehearsed by Davidson (1990, 302-305), according to which the correspondence theorist ought to be able to meet the challenge of *locating* 'the fact or part of reality' to which a true sentence corresponds, then this feature of Tarski's definition would prove its incompatibility with the correspondence theory of truth. However, the claim that sentences are made true or false by knowledge-independent facts or states of affairs to which they either correspond or fail to correspond is not part of our minimal characterization of correspondence truth, which only requires that sentences be made true or false by the nature of knowledge-independent *reality*. Commitment to correspondence truth does not involve commitment to an ontology of knowledge-independent facts or states of affairs[4].

[4] I believe that facts should be conceived as being, at least in part, linguistic constructs. In my opinion, facts supervene on reality *under a linguistic description*. If this is true, it will be plausible to suppose that 'correspondence' contexts fail to be transparent (the statement that Naples is farther north than Red Bluff does not 'correspond' to the same fact as the statement that the largest Italian city within thirty miles of Ischia is farther north than Red Bluff). Therefore, the Frege-Church ('slingshot') argument harnessed by Davidson (1984, 42) to prove that if true sentences correspond to facts, there is just one Great Fact to which they all correspond, cannot even get under way, and Davidson's (1990, 303) claim that 'there is nothing interesting or instructive to which true sentences might correspond' proves beside the point. One may very well say that true sentences correspond, and false sentences fail to correspond, to (language-dependent) facts, as long as one clearly recognizes that their success/failure to correspond to (language-dependent) facts is supervenient upon their success/failure to being satisfied by (language-independent) sequences of objects.

1.5 Correspondence Truth and Realism

What is then the relationship between (our minimal characterization of) correspondence truth and realism? Putnam once wrote that 'Whatever else realists say, they typically say that they believe in a Correspondence Theory of Truth' (Putnam 1976, 177). But in the light of our previous considerations, commitment to correspondence truth seems to *presuppose* commitment to metaphysical realism (recall that the third condition in our characterization of correspondence truth refers to the 'knowledge-independent nature' of reality!) rather than to *replace* it. Thus it seems to me altogether too quick to identify metaphysical realism with pure and simple commitment to correspondence truth, as many contemporary writers are inclined to do. For one thing, one has to take into account the disquotational theory of truth. I have said that I do not believe that the disquotational theory of truth represents a viable philosophical option, but I could be wrong. If indeed I am wrong, Loar's (1987, 83) argument that a disquotational theory is consistent with an evidence-transcendent notion of truth because the disquotational sense of 'true' is all that one needs in order to say that any statement in a given class might be true though unverifiable will suffice to refute a straightforward identification of metaphysical realism with correspondence truth.

But even if I am right to think that the disquotational theory of truth is unsuccessful, the connection between metaphysical realism and correspondence truth is mediated by epistemological considerations. The argument for this is given by Devitt (1991b, 44-45). Its formulation is in terms of the knowledge-independent existence of the *natural* world, because of course most people (and Devitt among them) will assume that if anything is likely to have a knowledge-independent existence at all, that is the natural world; but the point is absolutely general, and if someone has different ontological inclinations, the argument can easily be adapted to any ontology involving the existence of knowledge-independent entities.

Assuming that epistemic truth provides the only viable alternative to correspondence truth, Devitt argues, roughly, that one should expect to meet huge epistemological difficulties if one were to try to combine acceptance of metaphysical realism about the natural world with rejection of correspondence truth. Devitt calls his an 'abductive' argument from epistemic truth to anti-realism. It must be emphasized that the goal of this argument is not to provide a *reductio* of epistemic truth, but merely to

highlight the epistemological difficulties arising from the attempt to conjoin epistemic truth with metaphysical realism.

What is 'epistemic truth'? Epistemic theories of truth are sometimes identified with, sometimes contrasted to, 'pragmatic' theories. Those writers who contrast them, see the latter as involving a higher degree of idealization than the former. I shall use 'epistemic' as a general term for both sorts of theories. An epistemic theory of truth, then, is a theory which explains what it is for a sentence to be true or false in terms of some epistemic notion such as provability, verifiability, rational acceptability, warranted assertability, *et similia*. Since there appears to be an obvious gap between what *I* can prove and what can be proved, between what *I* can verify and what an observer in a better epistemic position than mine could verify, etc., epistemic theories of truth usually contain reference to some sort of *ideal* epistemic situation or cognitive procedure. For example, Peirce (1934, 394 [CP 5.407]) takes truth to be what the scientific community will eventually agree upon if its members carry on their research long enough[5], and Putnam (1983, 84) takes truth to be what would one would be justified to believe under epistemically ideal conditions.

Now acceptance of an epistemic theory of truth is obviously at odds with acceptance of a correspondence theory of truth. For it is highly doubtful if there can be any world where the relevant cognitive subjects have no super-human cognitive faculties in which these theories may turn out to be at least extensionally equivalent. Devitt's claim is that replacement of correspondence truth with epistemic truth will confront a realist with virtually insuperable epistemological difficulties.

According to Devitt, the closest link between the epistemic doctrines of truth and (metaphysical) realism is displayed by an application of Tarski's material requirement. Assuming that every acceptable definition of truth should have as a consequence all instances of the (T) schema, then, if 'T_E' stands for 'true as defined by the epistemic doctrine E', E will require that the appropriate instances of

---

[5] In other passages Peirce seems to subscribe to the correspondence theory of truth. But the pragmatist principle that the entire meaning of a conception lies in the sum of the practical consequences that follow from its truth and the related claim that there can be no truth that is in principle unknowable to rational inquirers appear to sit more comfortably with an epistemic theory of truth. In private conversation, Paul Forster has suggested to me that Peirce's goal was probably to show that the correspondence, coherence, epistemic and instrumentalist accounts of truth can be regarded, when purged of their metaphysical assumptions, as equivalent. However this may be, it is a fact that reference to the 'Peircean limit of inquiry' has become commonplace in discussions of the epistemic theory of truth.

S is T_E iff *p*

hold. For example:

'Caesar had five moles' is T_E iff Caesar had five moles

But why should such instances of our modified (T) schema hold? Devitt invites us to consider an extravagant version of epistemic theory of truth: T_E = is affirmed by the Pope. Why should

'Caesar had five moles' is affirmed by the Pope iff Caesar had five moles

hold? An explanation can only be provided by some account of the relationship between the Pope's assertions and the relevant states of the world. We could say, for example, that the Pope's assertions are invariably true because the world is created by the Pope's word. Or we could say that they are invariably true because the Pope enjoys some infallible insight of a divine origin into how the world is. These two accounts lead to different attitudes towards the knowledge-independent existence of the external world, respectively an anti-realist and a realist one. Therefore the doctrine that 'to be true is to be affirmed by the Pope' appears to have some bearing on the issue of (metaphysical) realism only *via* some further theory about the relationship between the Pope's assertions and the relevant states of the world. But while in the case of our papist doctrine we can imagine that some neo-scholastic philosopher may find the theory that leads to Realism more attractive than the theory that leads to a Pope-creator, in the case of the epistemic doctrines currently advanced in epistemological circles the situation is completely reversed:

> The problem which the epistemic doctrine poses for the Realist is that it is hard to find a plausible Realist epistemology [i.e., a plausible epistemology consistent with metaphysical realism] to do the explanatory job for most, if not all, the likely candidates to be T_E.
>
> (Devitt 1991b, 45)

If, for example, T_E were to be construed in terms of warranted assertability, the external world realist could hardly devise an explanation of the incredibly close link between a

state of the world, Caesar having five moles, and our being warranted in asserting 'Caesar had five moles', that the epistemic doctrine in question would commit her to. While in a suitably strong sense of 'warranted' it might seem sensible to assume that we may be warranted in asserting that 'Caesar had five moles' *only if* Caesar had five moles, it is much more difficult to explain why, if Caesar had five moles, we should be warranted in asserting that he did. For which epistemological doctrine assuming the knowledge-independence of the external world could ever guarantee the warranted assertability of *all* true statements? Doesn't the warranted assertability of 'Caesar had five moles' depend on the availability of relevant evidence? And isn't the availability of relevant evidence — even in the Peircean limit of inquiry — a largely contingent matter? If this is the case, the upholder of epistemic truth will be forced either to denounce such undecidable statements as 'Caesar had five moles' as meaningless (thus leaving such states of affairs as Caesar's having five moles out of the world[6]), or else to relax her criteria of warranted assertability in order to allow a decision about the truth-value of such statements even in the absence of (what we would usually regard as) satisfactory evidence. In both cases, the conclusion appears to be that commitment to epistemic truth will lead to a conception of the world as somewhat shaped and structured by our cognitive powers. (A third option might be to build the knowledge-independence of the [external] world into the very concept of epistemic truth by introducing an extremely high degree of idealization in its definition. One might resolve, in other words, to construe the notion of warrant as having only an extremely weak connection with actual human cognitive abilities. However, if one requires that an adequate definition of epistemic truth preserve such [realist] intuitions as that there must be a fact of the matter as to whether Caesar had five moles even if nobody will ever possess sufficient evidence to settle the question, the resulting conception of truth (i) will be hardly distinguishable from correspondence truth and (ii) will disappoint expectations that epistemic truth may prove epistemically more accessible than correspondence truth, thus undermining the main motivation for its introduction).

In sum, even if epistemic truth does not logically imply the denial of the knowledge-independent existence of the external world, it is nevertheless likely to lead

---

[6] Since I do not believe that states of affairs are part of the furniture of the world (see footnote 4 above), what I mean is, more precisely, that the upholder of epistemic truth will be forced to say that there is no fact of the matter as to whether Caesar had five moles or not.

to anti-realism for lack of viable epistemological explanations of why the instances of the (T) schema that follow from a particular choice of T_E should hold on the basis of a realist view of the external world. So it seems that acceptance of correspondence truth, though not identical with realism tout-court, is nevertheless crucial for its successful defence — unless, that is, one is willing to make do with disquotational truth, which is not the strategy I am investigating in this work.

Saying that acceptance of correspondence truth is *crucial* for a successful defence of realism is a weaker claim than saying that realism about the external world merely *is* commitment to correspondence truth for the interpretation of natural world discourse, or that it *essentially is* commitment to correspondence truth for the interpretation of natural world discourse. Though I believe that only the weaker claim is warranted, my argument in this work will be in fact a defence of metaphysical realism in general, and of metaphysical realism about the external world in particular, only in so far as it will be a defence of correspondence truth. Since I believe that the rejection of correspondence truth would seriously undermine the plausibility of metaphysical realism, I take it that a successful defence of correspondence truth vis-à-vis the epistemic doctrines will be of some value for the metaphysical realist. My hope is of course that my arguments so far have not been completely unsuccessful. However, it is not really important for what follows if they have not convinced the reader to relinquish her favoured Dummett-style view of the issue of realism. Nothing crucial hangs on that. If realism simply *is* commitment to correspondence truth, then the relevance of what follows to the issue of realism will be even more straightforward than I am prepared to claim.

The Problem of the Epistemic Accessibility of Truth

Acceptance of epistemic truth is often recommended on the supposition that commitment to correspondence truth is bound to make reality epistemically inaccessible and knowledge impossible. In this chapter I shall start to dispute the plausibility of this claim by arguing that if epistemic truth is understood in a reasonably idealized fashion, correspondence truth turns out to be at least as epistemically accessible as its rival. (In subsequent chapters I shall argue that besides being *at least as* accessible as epistemic truth, correspondence truth may also prove *sufficiently* accessible to our cognitive efforts).

Section 2.1 introduces the 'inaccessibility' argument that has led many philosophers to subscribe to some form of anti-realism. Section 2.2 presents Putnam's argument that it is necessarily false that we are all brains in a vat, while section 2.3 argues that Putnam's point about the 'preconditions' of reference does not entitle internalist philosophers to dismiss the 'Brains-in-a-Vat' hypothesis as meaningless and the sceptic's challenge as harmless. Section 2.4 rejects Putnam's claim that knowledge can be shown to be possible — without disputing the adequacy of sceptical standards — simply by substituting commitment to epistemic truth for commitment to correspondence truth. And section 2.5 concludes that, *ceteris paribus*, any epistemic theory of truth falling short of strict verificationism will fail to make truth epistemically more accessible than a correspondence theory can.

2.1 The 'Inaccessibility' Argument Against Correspondence Truth

Dummett (1976) has a famous argument which is claimed to show that no truth-conditional theory of meaning committed to the existence of evidence-transcendent truths will succeed in providing an adequate account of the propositional knowledge implicitly possessed by the speakers of some language L. This is the widely discussed *semantic* argument from the acquisition and manifestation of linguistic competence. I am prepared to grant that a comprehensive defence of realism and correspondence truth will have to meet Dummett's challenge and show the feasibility of a theory of meaning involving ascription of evidence-transcendent truth-conditions to a large number of sentences in our language. But in this work I shall not embark on such a project[1], since it is a major undertaking that would not allow me to get on to my topic.

The basic reason that prompts many anti-realist philosophers to substitute epistemic truth for correspondence truth is not semantic, but *epistemological*: it is the fear that understanding truth as correspondence to a knowledge-independent reality will turn it, in Davidson's (1990, 304) phrase, into 'something to which humans can never legitimately aspire'. This is an old story. Kant voiced a similar feeling in the so-called *Jäsche Logic*:

> Truth, it is said, consists in the agreement of cognition with its object. In consequence of this mere nominal explanation, my cognition, to count as true, is supposed to agree with its object. Now I can compare the object with my cognition, however, only *by cognizing it*. Hence my cognition is supposed to confirm itself, which is far short of being sufficient for truth. For since the object is outside me, the cognition in me, all I can ever pass judgment on is whether my cognition of the object agrees with my cognition of the object.
>
> (Kant 1992, 557-558)

It is not clear if Kant saw himself as giving up what he called the 'nominal' explanation of truth altogether. In the *Critique of Pure Reason*, for example, he seems to endorse the view that 'truth consists in the agreement of knowledge with the object' (see Kant 1933, 220; the passage on p. 97 is more ambiguous). However, Kant's 'transcendental idealism' is clearly incompatible with any notion of truth that could satisfy our third

[1] The outline of a quasi-holistic theory of understanding which meets Dummett's challenge by showing how grasp of evidence-transcendent truth-conditions may be manifested in a person's behaviour can be found in Loar (1987).

requirement for correspondence truth. This is because Kant is adamant that the phenomenal world (that is, the only world he takes to be epistemically accessible to human beings), being structured by the categories of understanding, cannot be seen as existing independently of us and of our sensibility, and this of course means that it cannot be counted as knowledge-independent. Moreover, Kant seems to suggest that an adequate concept of truth must be a decidable one:

> The question here is, namely, whether and to what extent there is a criterion of truth that is certain, universal, and useful in application. For this is what the question, *What is Truth?*, ought to mean.
>
> (Kant 1992, 558)

Passages like this suggest that Putnam (1981, 60-64) could be right in presenting his own perspective as a development of Kant's transcendental philosophy. However that may be, old and new critics of correspondence truth seem mostly prompted by the desire to bridge the epistemic gap between evidence and truth they take as a consequence of the correspondence theory of truth.

We saw in section 1.4 that the correspondence theory of truth does not entail, by itself, any claim at all about the epistemic accessibility or inaccessibility of truth. There can be no doubt, however, that correspondence truth makes an epistemic gap between evidence and truth at least *possible*. For it is a consequence of the correspondence theory of truth that a statement's being true (false) is conceptually independent from its being *recognizably* true (false). For any *bona fide* descriptive statement *p*, the nature of knowledge-independent reality will suffice to make *p* either true or false even if no cognitive subject should ever be able to determine p's truth-value. But while upholders of correspondence truth address the problem of that gap only after their notion of truth is already in place, upholders of epistemic doctrines build the solution to the problem into their definition of truth from the outset.

The Kantian argument that, if truth is correspondence to a knowledge-independent reality, the object as it is in itself (*Ding-an-sich*) will forever transcend our cognitive grasp because we can never step out of the circle of our representations to compare them with the way the world is independently of our knowledge, is now usually rehearsed as an argument from the theory-ladenness and/or conceptual relativity of all our experience of the world. This is because various sorts of foundationalist epistemologies can be

construed as rehabilitating the idea of a direct comparison of our representations with their intended objects, in virtue of the claim that sense-data or even more substantial forms of direct experience ensure that immediate access to knowledge-independent reality that is commonly thought to be the *condicio sine qua non* for assessing the correspondence of our representations to the way the world is. Over and against this foundationalist claim (and the project of reconstructing all reality out of the unshakable foundations provided by sense-data or direct experience), the 'inaccessibility' argument against correspondence truth is now usually spelled out as the claim that the theory-ladenness and/or conceptual relativity of all our experience entails the transcendence of correspondence truth with respect to all evidence (in principle) available to human beings.

The 'inaccessibility' argument against correspondence truth (and metaphysical realism) is formulated or taken for granted in the writings of Richard Rorty, Hilary Putnam, Nelson Goodman, Arthur Fine, Brian Ellis, Nicholas Rescher, David Bloor, and many others[2]. It boils down to the following. To ascertain that a given statement is correspondence-true, one ought to have access both to one's own representations of the world *and* to the world as it is in itself, before any conceptualization, in order to compare their features and determine if they fit with each other. However, after Wittgenstein's private language argument and Sellars' critique of the 'given', it is widely agreed that our experience of the world is always theory-laden, and that its content is always shaped, to some extent, by our conceptual choices. Accordingly, if experience never grants us access to a knowledge-independent reality, we shall never be in a position to compare our representations of the world with the world as it is in itself, and correspondence truth will forever transcend all the evidence to which we can have access. If nevertheless we believe that we *do* currently assess our representations of the world with respect to their (likeness to the) truth and are therefore not prepared to yield to pure and simple scepticism, this result will force us to give up the claim that correspondence truth is the aim (or part of the aim) of our cognitive efforts.

---

[2] Davidson (1986) has a similar argument, but directs it against foundationalism rather than correspondence truth. Davidson's coherence theory of truth is to be construed as a 'criterial' theory (see Haack 1978, 88 ff. on the opposition between 'criterial' and 'definitional' theories of truth) if it is to be consistent with the correspondence theory that he also endorses in that paper. It is worth noting, though, that Davidson has recently retracted his (1986) commitment to correspondence truth (see Davidson 1990, esp. 303-304).

The 'inaccessibility' argument against correspondence truth provides what seems to me the strongest and commonest motivation for the development of different — especially epistemic — accounts of truth. This is the reason why my discussion will focus on those epistemic accounts of truth which are explicitly defended by their proponents on epistemological, rather than semantic or analytical, grounds. My discussion will not affect, for example, Nicholas Jardine's account, which is motivated by the alleged difficulty of devising a substantive conception of the 'relation of correspondence between beliefs and the world' (Jardine 1986, 11) and which is clearly not intended to make epistemic truth more accessible than correspondence truth. And as far as the 'inaccessibility' argument is concerned, I shall question neither the claim that our experience of the world is always theory-laden and/or shaped by our conceptual choices, nor the claim that whether a given statement about the world is correspondence-true will typically transcend all evidence (in principle) available to human beings. What I want to dispute is rather the suggestion that if these claims are true, correspondence truth will have to be epistemically inaccessible. I shall argue that the fact that a God's eye view is beyond the reach of human beings is not a good reason for giving up the idea that truth is correspondence to a knowledge-independent reality.

2.2 Putnam's 'Internalist' Refutation of the Sceptic

A dramatized version of the 'inaccessibility' argument against correspondence truth can be found in Putnam (1981, chs. 1-3)[3]. There Putnam proposes to distinguish 'two philosophical perspectives', which he calls the 'externalist' and the 'internalist' perspective. The externalist perspective is, in our terminology, the conjunction of metaphysical realism, correspondence truth, and the thesis that there is exactly one true and complete description of 'the way the world is'. Putnam calls this composite doctrine the *externalist* perspective precisely because 'its favorite point of view is a God's Eye point of view' (Putnam 1981, 49). On the other hand, an *internalist* perspective is described by him as dismissing any talk of reality implying an 'external' point of view:

[3] Putnam (1983, 1-25) contains a model-theoretic argument allegedly showing the impossibility of unambiguous reference which, if sound, would have disruptive consequences for the correspondence theory of truth. Some questionable premises of this argument, which I am not going to discuss here, are pointed out by Hacking (1983, 105-108).

> There is no God's Eye Point of View that we can know or usefully imagine, there are only the various points of view of actual persons reflecting various interests and purposes that their descriptions and theories subserve.
>
> (Putnam 1981, 50)

According to an internalist perspective, asking, What objects does the world consist of? only makes sense *within* a theory or description: prior to our conceptual choices, there is no ready-made world about which this question can meaningfully be asked. Putnam takes this to entail that acceptance of an internalist perspective will imply commitment to an epistemic theory of truth.

It is not clear to me whether a (metaphysical) realist should be committed, as Putnam believes, to the 'One True Theory' claim. Even when metaphysical realism is conjoined with a correspondence theory of truth, it is by no means obvious that the 'One True Theory' claim must follow, at least if the correspondence between language and world is not taken (as I argued it ought not to be taken) to be an isomorphism between sentences and knowledge-independent facts or states of affairs (on this issue, see Field 1982, 553 f.). Moreover, we saw in section 1.1 that a metaphysical realist can deny that the domain of what there is is coextensive with the domain of what we can think about or conceive of, thus denying that there may be even *one* true and complete description of the way the world is. However, even if Putnam presents his argument as a refutation of the externalist perspective, the 'One true theory' claim does not play any actual role in it. As far as Putnam's argument is concerned, the distinction between the internalist and the externalist perspective boils down to the distinction between the conjunction of metaphysical anti-realism and epistemic truth on the one hand, and the conjunction of metaphysical realism and correspondence truth on the other. Putnam's argument for the internalist perspective can thus be regarded as an argument for epistemic truth (and metaphysical anti-realism) *via* the rejection of correspondence truth.

Putnam asks us to imagine ourselves in the following situation (Putnam 1981, 5 ff.): our brain has been removed from our body and placed in a vat of nutrients to keep it alive. Our afferent nerve-endings have been connected to a super-scientific computer which causes us to have the illusion that everything is perfectly normal. If we try to raise our hand, the feedback from the computer will cause us to 'see' and 'feel' the hand being raised. And all the 'experiences' we are now having are nothing but the result of the computer's action on our nerve-endings.

This thought-experiment is in fact quite common in epistemology and the philosophy of mind, but Putnam's version of it makes the further assumption that it is not just you and me who are the victims of some evil (or benevolent) super-psychologist, but *all* human beings are in fact nothing but brains in a vat:

> Perhaps there is no evil scientist, perhaps [...] the universe just happens to consist of automatic machinery tending a vat full of brains and nervous systems. (Putnam, 1981, 6)

This assumption may seem overwhelmingly implausible, but Putnam maintains that such a state of affairs is not physically impossible (by this he means that it can described in a way which is consistent with the laws of physics). So he goes on to ask: 'Could we, if we were brains in a vat in this way, *say* or *think* that we were?' (Putnam 1981, 7). His answer his a resounding 'No'. The reason has to do with what Putnam takes to be the 'preconditions' for thinking about, representing, referring to things in the world, which he claims are not fulfilled in the case of the brains in a vat.

Putnam ascribes to Descartes and Locke the view that the reference of our mental representations is fixed by their similarity to the relevant objects in the world. This view he claims to be essentially flawed, because everything being similar to everything else in infinitely many respects, the similarity-relation cannot guarantee any *intrinsic* connection between our mental representations and their referents. His claim is in fact that 'one cannot refer to certain kinds of things, e.g. *trees*, if one has no causal interaction at all with them, or with things in terms of which they can be described' (Putnam 191, 16 f.). To think that our mental representations could have an *intrinsic* connection with their referents independently of any (direct or indirect) causal interaction would mean to accept, in Putnam's words, a 'magical theory of reference'. Sometimes Putnam seems to suggest that that one cannot reject such a theory other than by subscribing to a thoroughly causal theory of reference, but Wright (1992, 71) correctly notes that Putnam's argument merely requires the weaker assumption that reference is, in appropriate cases, a '*causally constrained* relation'.

If reference, then, presupposes a (direct or indirect) causal interaction with the object, what would be the referents of our words and mental representations if for some kind of cosmic accident all of us were brains in a vat? Putnam suggests three possible answers. If for some kind of cosmic accident all of us were brains in a vat, the word

'tree' might refer 'to trees in the image, or to the electronic impulses that cause tree experiences, or to the features of the program that are responsible for those electronic impulses' (Putnam 1981, 14). What the word 'tree' could not possibly refer to are actual trees (if any) outside the computerized image, for no causal connection at all would obtain, *ex hypothesis*, between them and the brains' usage of the word 'tree'.

> It follows that if their [of the brains] 'possible world' is really the actual one, and we are really the brains in a vat, then what we now mean by 'we are brains in a vat' is that *we are brains in a vat in the image* or something of that kind (if we mean anything at all). But part of the hypothesis that we are brains in a vat is that we aren't brains in a vat in the image (i.e. what we are 'hallucinating' isn't that we are brains in a vat). So, if we are brains in a vat, then the sentence 'We are brains in a vat' says something false (if it says anything). In short, if we are brains in a vat, then 'We are brains in a vat' is false. So it is (necessarily) false.
>
> (Putnam 1981, 15).

This argument is claimed to provide a 'transcendental' (i.e., based on the 'preconditions' of our capacity to refer to things in the world) refutation of the 'sceptical' hypothesis that we might be all brains in a vat:

> The existence of a 'physically possible world' in which we are brains in a vat [...] does not mean that we might really, actually, possibly *be* brains in a vat. What rules out this possibility is not physics but *philosophy*.
>
> (Putnam 1981, 15)

The deeper aim of the argument, though, is to highlight the epistemological unattractiveness of the externalist perspective (= metaphysical realism + correspondence truth) by showing that it is committed to take seriously some wildly evidence-transcendent 'sceptical' hypotheses which could be easily dismissed as meaningless from the internalist perspective (= metaphysical anti-realism + epistemic truth):

> Internalist philosophers dismiss the 'Brain in a Vat' hypothesis. For us, the 'Brain in a Vat World' is only a *story*, a mere linguistic construction, and not a possible world at all. [...] For the externalist philosopher, on the other hand, the hypothesis that we are all Brains in a Vat cannot be dismissed so simply. For the truth of a theory does not consist in its fitting the world as the world presents itself to some observer or observers [...] but in its corresponding to the world as it is in itself.
>
> (Putnam 1981, 50)

And of course, the suggestion is that if the hypothesis that we are all brains in a vat has to be taken seriously, no knowledge of the 'external' world will be possible unless some proof of the non-actuality of that possibility can be developed. In other terms, the claim is that while the identification of the 'preconditions' of reference enables internalist philosophers to dismiss the 'Brain in a Vat' hypothesis as meaningless, externalist philosophers are compelled by their own conception of truth to regard it as a genuine possibility even if its realization would in fact prevent us from entertaining it[4]. But this means that externalist philosophers unwilling to embrace scepticism will have to face the exceedingly difficult task of ruling out the actuality of the 'Brain in a Vat' hypothesis and of all those 'sceptical' worlds which are deliberately conceived so as to be evidentially indistinguishable and yet utterly different from (what we take to be) the actual world. (Here Putnam is rehearsing a familiar sceptical argument, to which I shall return in chapter 4. This is why I said, at the beginning of this section, that the 'Brain in a Vat' hypothesis can be seen as providing a dramatized version of the inaccessibility argument against correspondence truth. While the inaccessibility argument relies on general considerations about the theory-ladenness and conceptual-dependence of our experience of the world, the 'Brain in a vat' argument presupposes more substantive assumptions about the nature and function of human brains and nervous systems. But the point is just the same: correspondence truth is hopelessly evidence-transcendent. Given the possibility of sceptical alternatives to the actual world, human beings will never be in a position to tell whether their beliefs are correspondence-true or not, because their evidence will always underdetermine the content of their judgements).

---

[4] The fact that externalist philosophers are not committed to deny the claim that the brains in Putnam's story cannot conceptualize their predicament is relevant to whether his argument that we are not brains in a vat may be taken to provide not just a ground for being dissatisfied with the externalist perspective, but a genuine refutation of it. Wright (1992, 90-94) notices that if the externalist perspective is committed to claiming the irrefutability of the hypothesis that the real world could be dramatically different to the way we take it to be on the basis of our best theories, then the availability of a proof that we are not brains in a vat might seem to provide a refutation of externalism. But apart from the fact that it is far from clear that the conjunction of metaphysical realism with correspondence truth involves the *irrefutability* of sceptical hypotheses, Wright correctly argues that Putnam's rejection of the hypothesis that we are all brains in a vat fails to provide a refutation of the externalist perspective precisely because this is not commited to claim that, if we were all brains in a vat, we would be able to conceptualize our predicament. I shall return to this issue in the next section; in the meantime, it is fair to acknowledge that Putnam himself does not claim that his argument provides a conclusive refutation of the externalist perspective.

2.3 A Criticism of Putnam's Refutation of the Sceptic

Putnam's contention that the utterances of the brains in the 'Vat World' cannot be interpreted as saying what the realist would like them to be saying has been the subject of extensive debate in the eighties. In this section I shall not try to reject this contention. Rather, I shall question Putnam's claim that his point about the 'preconditions' of reference enables internalist philosophers to dismiss the 'Brain in a Vat' hypothesis as meaningless and the sceptic's challenge as harmless.

Putnam's argument can be rendered as follows:

1	(1)	One cannot refer to certain kinds of things, e.g. *trees*, if one has no causal interaction at all with them, or with things in terms of which they can be described [reference is a causally constrained relation; compare Putnam 1981, 16-17]
1	(2)	'Brain' cannot in Vat-English designate brains, but only brains-in-the-image, and 'vat' cannot in Vat-English designate vats, but only vats-in-the-image [from 1; compare Putnam 1981, 14]
1	(3)	If we are brains in a vat, then the sentence 'We are brains in a vat' says something false [from (2); compare Putnam 1981, 15]
1	(4)	The sentence 'We are brains in a vat' is (necessarily) false [from (3); compare Putnam 1981, 15]

This argument is claimed to prove that the sentence 'We are brains in a vat' is necessarily false *in the sense that the supposition that it is entertained or enunciated implies its falsity* (see Putnam 1981, 7-8). But does the fact that the sentence 'We are brains in a vat' is necessarily false in *this* sense amount to a refutation of the sceptic? Does Putnam's argument prove *that we are not brains in a vat*? Here is how Putnam's argument must be tidied up in order to prove not only that the sentence 'We are brains in a vat' is (necessarily) false, but also that we are not brains in a vat:

1	(1)	One cannot refer to certain kinds of things, e.g. *trees*, if one has no causal interaction at all with them, or with things in terms of which they can be described [reference is a causally constrained relation]
1	(2)	'Brain' cannot in Vat-English designate brains, but only brains-in-the-image, and 'vat' cannot in Vat-English designate vats, but only vats-in-the-image [from 1]

1	(3)	If we are brains in a vat, then the sentence 'We are brains in a vat' says something false [from (2)]
4	(4)	If the sentence 'We are brains in a vat' says something false, then we are not brains in a vat [by disquotation]
1, 4	(5)	If we are brains in a vat, then we aren't [from (3) and (4)]
1, 4	(6)	We are not brains in a vat [from (5)]

Putnam's argument now leads to the conclusion that we are not brains in a vat. But is it valid? And does it prove that the sceptic is wrong?

Putnam's argument is expressed in English, in the sense that any language which turned out to be Putnam's language would be English. Let us consider (2) first. (2) is in fact the conjunction of

(2') The set of brains is disjoint from the set of things designated by 'brain' in Vat-English and the set of vats is disjoint from the set of things designated by 'vat' in Vat-English

and

(2'') The set of brains-in-the-image is identical with the set of things designated by 'brain' in Vat-English and the set of vats-in-the-image is identical with the set of things designated by 'vat' in Vat-English.

In a discussion of Putnam's argument, Kinghan (1986) points out that the sentence

(*) The set of brains is identical with the set of things designated by 'brain' in English and the set of vats is identical with the set of things designated by 'vat' in English

is, as a sentence of English, necessarily true (by disquotation). Kinghan goes on to point out that if Vat World were our world and Vat-English were English, then the sentence

(**) The set of brains is identical with the set of things designated by 'brain' in Vat-English and the set of vats is identical with the set of things designated by 'vat' in Vat-English

would also have to be true. Therefore if Vat World *is* our world and Vat-English *is* English, then (2') must be false. By assuming that (2') is entailed by (1), Putnam is in fact presupposing that Vat-English and English are two different languages, i.e., that we are not brains in a vat. From this Kinghan concludes that Putnam's argument is viciously circular. But could not Putnam's assumption that (2') is entailed by (1) be just an harmless oversight? Is (2') really essential to the derivation of the conclusion of his argument?

The assumption that Vat-English and English are two different languages may be argued to be unnecessary to the derivation of (3), because (3) follows from (2'') alone. (2''), however, is not itself above suspicion. For what follows from (1) is that brains-in-the-image and vats-in-the-image are among the *potential* referents of 'brain' and 'vat' in Vat-English, not that they are their *actual* referents. To validly derive (2'') from (1) one ought to argue, among other things, that 'real' brains and vats cannot be described in terms of brains-in-the-image and vats-in-the-image.

But Putnam's argument can be shown to be invalid even if it is granted that (2'') can be validly derived from the conjunction of (1) and some further, plausible assumptions. As we do not know in advance whether we are brains in a vat or not, there are two cases we have to consider:

A. *We are not brains in a vat and Vat-English and English are two different languages*[5]. In this case, steps (1) to (3) can, with the above-mentioned provisos, be granted. However, if Vat-English and English are two different languages, it is step (4) that fails to be warranted. For the sentence mentioned in the antecedent of (4) is a sentence in Vat-English, while the consequent of (4) is a sentence in English. Therefore the truth of the resulting conditional does not follow by disquotation. Putnam's argument thus fails to establish the conclusion that we are not brains in a vat.

B. *We are brains in a vat and Vat-English and English are one and the same language.* In this case, (**) tells us that the occurrences of 'brains' and 'vat' in the antecedent of (3) designate the same sets of things as the occurrences of 'brains' and 'vat' in tokens of the sentence-type mentioned in its consequent. This makes it very difficult to see how, if we are brains in a vat, the sentence 'We are brains in a vat' could

[5] It goes without saying that, if we are not brains in a vat, then, necessarily, we are not brains in a vat. The point of distinguishing cases (A) and (B) is to see whether Putnam's argument is *valid*, that is, if it proves its conclusion both if we are brains in a vat and if we aren't.

fail to be true. For if the occurrences of 'brains' and 'vat' in the antecedent of (3) designate the same sets of things as the occurrences of 'brains' and 'vat' in tokens of the sentence-type mentioned in its consequent, then '"We are brains in a vat" is true' follows by disquotation from the assumption that we are brains in a vat. Indeed, (3) can be true only if Vat-English and English are two different languages, but then, as we have seen, step (4) is unwarranted and Putnam's argument cannot establish the conclusion that we are not brains in a vat.

As the two cases I have discussed are logically exhaustive, I conclude that Putnam's argument fails to provide the purported refutation of the sceptic. This is so even if Putnam's claim that the sentence 'We are brains in a vat' is necessarily false (in the sense that the supposition that it is entertained or enunciated implies its falsity) does follow from the fact that it is false both if those who utter it are not brains in a vat (because they aren't) and if they are (if it is granted that brains in a vat cannot refer to real brains and vats, but only to brains-in-the-image and vats-in-the-image).

This is totally unsurprising. From the fact that there is no possible world in which I can truly utter the sentence 'I do not exist', it does not follow that there is no possible world in which G.V. does not exist. By the same token, from the fact that there is no possible world in which we can truly utter the sentence 'We are brains in a vat', it does not follow that there is no possible world in which we are brains in a vat. And this is enough for the sceptic to carry off the victory. For the sceptic does not claim that we *are* brains in a vat (if she did, she would be committed to *justify* the claim that we are). She merely points out that, for all we know, we *might* be brains in a vat, from which she draws the conclusion that we do not know all the familiar things we believe we know. The fact that, if we were brains in a vat, we could not truly say that we were, is therefore irrelevant to her claim that there are possible alternatives to (what we take to be) the actual world that cannot be ruled out on the basis of the evidence to which we have access.

The import of the claim that the sentence 'We are brains in a vat' is necessarily false is thus much narrower than Putnam would like to maintain. What can be salvaged of Putnam's argument that we are not brains in a vat is in fact an argument about the expressive powers of the inhabitants of Vat World, on the assumption that their world is not our world (and that reference is a causally constrained relation, and that brains-in-

the-image and vats-in-the-image are not the sort of things in terms of which we could describe real brains and vats):

1	(1)	One cannot refer to certain kinds of things, e.g. *trees*, if one has no causal interaction at all with them, or with things in terms of which they can be described [reference is a causally constrained relation]
2	(2)	We are not brains in a vat [Assumption]
1, 2	(3)	'Brain' cannot in Vat-English designate brains, but only brains-in-the-image, and 'vat' cannot in Vat-English designate vats, but only vats-in-the-image [from (1) and (2)]
1, 2	(4)	What we now mean by uttering the sentence 'We are brains in a vat' cannot be expressed in Vat-English [from (2) and (3)]
1	(5)	If we are not brains in a vat, then what we now mean by uttering the sentence 'We are brains in a vat' cannot be expressed in Vat-English [from (2) and (4)]

(5) represents, I think, the strongest result that follows from Putnam's argument that the sentence 'We are brains in a vat' is necessarily false. But if Putnam's argument cannot be construed as a proof that we are not brains in a vat, then not only will it fail to provide a refutation of the sceptic, it will fail to provide the desired argument against the externalist perspective Putnam is in the business of rejecting.

2.4 Could *I* Be a Brain in a Vat?

Even if Putnam's thought-experiment cannot be seen as providing a refutation of the sceptic, still one could claim that it does lend some support to the internalist perspective, in so far as it draws our attention to the epistemological advantages enjoyed by that perspective as compared to its externalist alternative. This is because Putnam's thought-experiment describes a situation in which epistemic truth appears incomparably more accessible than correspondence truth — at least if one shares Putnam's view of the referential powers of the brains in Vat World. Those brains appear to have all the evidence they need to decide whether the sentence 'We are all brains in a vat' is, as they pronounce it, epistemically true. If epistemic truth is the only kind of truth one ought

to take pains to consider, then the fact that they cannot even conceive what the world should be like for the English sentence 'We are all brains in a vat' to be correspondence-true will cease to appear as a serious impairment of their cognitive abilities. And one could think this to be enough to turn the epistemological balance in favour of epistemic truth and against correspondence truth.

I am not convinced by this line of reasoning. In this section I shall argue that if epistemic truth is understood, as Putnam himself declares to understand it, as an idealization of rational acceptability, then, unless Vat World *is* our world (which of course Putnam does not want to prove), there is nothing in his thought-experiment that justifies the conclusion that epistemic truth is more accessible to human beings than correspondence truth.

We saw in the last chapter that upholders of epistemic theories of truth are apt to employ the concept of an 'ideal epistemic situation' in their definitions of truth. On the one hand they will argue, like Putnam, that to define truth as a correspondence between our representations and the world as it is in itself is to define truth 'from the God's Eye Point of View'. On the other hand, they will usually admit that truth cannot be straightforwardly identified with provability, verifiability, rational acceptability, or warranted assertability in *real* epistemic situations. For real epistemic situations are often not good enough to prove or verify a statement, or to make it rationally acceptable or warrantedly assertible. For example, a sentence *p* may be, say, verifiable for S at time t_0, but unverifiable for S at every time t_i later than t_0[6]. Shall we say that truth can be 'lost' and that *p* can be true at t_0 and false at every time t_i later than t_0? Or shall we say that a sentence *p* is true at any time if it is true at any one time? The first option seems absurd. But the second one has the unpalatable consequence that the truth of a sentence turns out to be a very contingent matter. S might have been sleeping at t_0. Or she might have been in a different place. In many such cases, *p* might have been as unverifiable for S at t_0 as at any other time, and *p* might fail to be true for what is apparently a very contingent reason. Accordingly, upholders of epistemic truth are likely to regard the concept of an 'ideal epistemic situation' as providing a third way between defining truth with reference to a real epistemic situation and defining truth with reference to *God's* epistemic situation. Here are two quotes from Putnam:

---

[6] It is not essential that S should be an individual. S could very well be the scientific community or, for that matter, the whole human kind.

> We speak as if there were such things as epistemically ideal conditions, and we call a statement 'true' if it would be justified under such conditions.
> (Putnam 1981, 55)

> The suggestion I am making, in short, is that a statement is true of a situation just in case [...] a sufficiently well placed speaker who used the words in that way would be fully warranted in *counting* the statement as true of that situation.
> (Putnam 1988, 115)

There are clues of a shift, in Putnam's writings, as far as the issue of the attainability of an 'ideal epistemic situation' is concerned. Sometimes Putnam seems to hold that no ideal epistemic situation can ever be a 'real' epistemic situation. In more recent writings, however, he takes a more optimistic attitude towards the attainability of epistemically ideal conditions, as though a number of real situations might be epistemically 'ideal'. The consequences of these different interpretations of the concept of an 'ideal epistemic situation' will emerge in the course of the discussion.

My point is that while the concept of an ideal epistemic situation does lend some plausibility to the epistemic conception of truth, it deprives it of all the anti-sceptical appeal characteristic of its most strictly verificationist versions. Not that I regard the elimination of sceptical alternatives to the actual world as a precondition for having any empirical knowledge at all. This is Putnam's belief, not mine. But his claim that his conception of truth as idealized rational acceptability does eliminate such alternatives seems to me seriously undersupported. Granted, if being-a-brain-in-a-vat is merely being-a-brain-in-a-vat-in-the-image, i.e. *verifiably* being-a-brain-in-a-vat, then we need not bother about evidence-transcendent sceptical hypotheses, for they are simply meaningless. But as soon as reference to an ideal epistemic situation is built into the definition of epistemic truth, this turns out to be epistemically on a par, as far as its accessibility is concerned, with correspondence truth.

My argumentative strategy will be to force the supporter of epistemic truth to commit herself to some specific judgment about what she would be prepared to count as an ideal epistemic situation by confronting her with what apparently *is* an ideal epistemic situation.

I am awake and in good physical health, I am not under the effects of drugs or alcohol, I am sitting in a comfortable position and light is good, I came on foot by Southampton Row and Kingsway, etc. As J.L. Austin's plain man would say, 'Well, if that's not an ideal epistemic situation for the purpose of knowing that I am in the LSE

library writing the second chapter of my Ph.D. thesis, *I don't know what is*'. Granted, Austin's (1962) plain man would never write a thesis about the epistemic accessibility of truth, but we can perhaps pass over that and ask the following question: Will the upholder of epistemic truth agree with (our caricature of) Austin's plain man?

If she doesn't, it seems to me that no conceivable situation will ever qualify, to her eyes, as an *ideal* epistemic situation. But if no situation accessible to human beings will ever qualify as ideal, epistemic truth will be as inaccessible as correspondence truth allegedly is.

If on the contrary the upholder of epistemic truth is willing to describe this situation as an *ideal* epistemic situation, I ask her to suppose that last night some super-psychologists got hold of me while I was asleep, took me to their laboratory in the countryside, removed my brain from my body, and transformed me in a brain in a vat, which (who) now believes 'I am in the LSE library writing the second chapter of my Ph.D. thesis'.

If I am a brain in a vat in a sceptical world, it is *false* that I am in the LSE library writing the second chapter of my thesis. But on an epistemic account of truth, that can be false only if my epistemic situation is *not* an ideal epistemic situation. This is because my epistemic situation as a brain in a vat (call it *BV*) is *ex hypothesis* evidentially indistinguishable, from my perspective, from my epistemic situation as a research student in the LSE library (call it *RS*); so it is rational for me to believe that I am in the LSE library writing the second chapter of my Ph.D. thesis.

Now I have to appeal to the distinction between 'internalist' and 'externalist' views of epistemic justification. These are standard terms in recent epistemology and therefore I will not replace them in order to avoid confusion with Putnam's own distinction between the internalist and the externalist 'perspectives'. But I must warn the reader that these two distinctions are far from equivalent (more about this in section 3.1). An *internalist* view of epistemic justification requires that a person have 'cognitive grasp' of whatever makes her beliefs justified (in other terms, only what is *within* that person's 'perspective' is allowed to determine the justifiedness of her beliefs). On the contrary, an *externalist* view of epistemic justification is such that a belief may be (as a matter of fact) epistemically justified even if the subject has no reason for thinking that it is (we shall have to introduce some refinements to this terminology in the next chapter, but so much will do for our present purposes).

If we take an internalist view of epistemic justification, *BV* and *RS* are epistemically equivalent, that is, no belief which is justified in *RS* can fail to be justified in *BV*, and no belief which is justified in *BV* can fail to be justified in *RS*. As the set of the beliefs which are justified in *BV* and the set of the beliefs which are justified in *RS* are extensionally equivalent, if it is false that I am in the LSE library writing the second chapter of my Ph.D. thesis, not only must *BV* be non-ideal, but *RS* must be non-ideal as well. But then, it appears that no epistemic situation accessible to human beings will ever qualify as an ideal epistemic situation, and epistemic truth will be as inaccessible as correspondence truth allegedly is.

If, on the contrary, we take an externalist view of epistemic justification, *RS* can qualify as ideal while *BV* can fail to do so (because what tells them apart need not be accessible to my subjective 'perspective'). But then, since I cannot possibly tell which of these situations is *my* situation, I cannot tell whether I am in an *ideal* epistemic situation or not. Therefore epistemic truth appears to be as inaccessible as correspondence truth allegedly is.

However, there is a different argumentative line that someone like Putnam may wish to take. One can argue that if I am a brain in a vat in the countryside, it is *true* that I am in the LSE library writing the second chapter of my Ph.D. thesis. One can say that the 'brain in a vat' hypothesis is a mere story, that in the world to which I have cognitive access it is true that I am in the LSE library writing the second chapter of my Ph.D. thesis, and that it makes no sense to imagine that there could be a further 'world' lying behind the world in which I live and evidentially indistinguishable from it. In other terms, *BV* is as ideal an epistemic situation as *RS*, because they are in fact one and the same situation.

This line of argument can be rejected as blatantly *ad hoc*. For in the case we are now discussing the upholder of epistemic truth cannot take refuge in the claim that reference is a causally constrained relation to justify her claim that if I am a brain in a vat in the countryside, then I cannot talk about things in the real world. Saying that, if I am a brain in a vat in the countryside, the sentence 'I am in the LSE library writing the second chapter of my Ph.D. thesis' is made false by the 'real' world, does not involve any commitment to a magical theory of reference. Putnam cannot argue that if I am a brain in a vat I cannot refer to the 'real' LSE library and to the 'real' draft of the second chapter of my thesis *because* my use of the phrases 'LSE library' and 'second

chapter of my Ph.D. thesis' fails to bear any relevant connection to the 'real' library and to the 'real' draft. For I learnt to use the phrases 'LSE library' and 'second chapter of my Ph.D. thesis' in the real world, and therefore there is nothing 'magic' about my ability to refer to the real LSE library and to the real draft[7].

Of course the fact that there is nothing 'magic' about my ability to refer, as a brain in a vat in the countryside, to things in the real world only exposes the *ad-hoc*ness of the attempt to win support for epistemic truth by claiming that the operation I underwent robbed me of the possibility of referring to the real world. The fact that reference is a causally constrained relation does not justify the claim that my being a brain in a vat in the countryside prevents me from referring to things in the real world, but I have no proof that if I am a brain in a vat in the countryside, then the words I learnt to use before the operation *do* refer to things in the real world. However, it seems much more sensible to suppose that the operation would cause me to be deluded about my present whereabouts than that it would change the referents and truth-values of a large number of my beliefs about the past. For example, I may now believe that two days ago I was in the LSE library reading a copy of Ronald Giere's book *Explaining Science*. If I am a brain in a vat in the countryside, this belief and the belief that I now am in the (same) LSE library writing the second chapter of my Ph.D. thesis cannot both be true (if they could, one would have to suppose that tokens of the phrase 'LSE library' could refer sometimes to 'real' and sometimes to 'virtual' buildings — which of course would contradict the assumption that if a I am a brain in a vat in the countryside I can *never* talk about real things). But then, it seems much more plausible to suppose that it is the latter belief that is false than that the operation instantaneously deprived me of the capacity of talking about the real world by miraculously changing the referents of the phrases 'LSE library' and 'Giere's book *Explaining Science*' in the sentence 'Two days ago I was in the LSE library reading a copy of Ronald Giere's book *Explaining Science*'. After all, the super-psychologists could resolve to undo the effects of their operation and give my body back to me. If they tell me what I went through, the most natural way of

---

[7] Putnam (1981, 16) himself admits that 'If the Brains in a Vat will have causal connection with, say, trees *in the future*, then perhaps they can *now* refer to trees by the description "the things I will refer to as 'trees' at such-and-such a future time"'. But the situation we are now discussing is of corse perfectly symmetric.

The inapplicability of Putnam's 'magical theory of reference' argument to those versions of the brains-in-a-vat story which do not feature the Cosmic Accident, but the Evil (or Benevolent) Scientist, is pointed out also in Wright (1992, 81-84).

framing a causally coherent account of my experience will not be by talking of miracolous transitions among different, totally unrelated worlds, but by contrasting genuine and merely delusory perceptual states within a unique world.

This latter account is supported by the fact that *BV* does not seem to be the sort of possibility we should want to dispose of a priori on purely analytical or metaphysical grounds — not even by saying that it is in fact the same situation as *RS*. For it is not part of the hypothesis (i) that no conceivable evidence could *ever* enable me to tell *BV* from *RS*; and (ii) that *BV* and *RS* are evidentially indistinguishable *for everybody* (surely they are not evidentially indistinguishable for the super-psychologists). We may want to dismiss a priori Putnam's hypothesis that we might be *all* brains in a vat without any chance ever to trespass the boundaries of our Vat World, but surely we will not want to dismiss a priori the hypothesis that *I* might be the victim of evil super-psychologists, if I stand any chance of finding out what other people from a different but definitely not Godlike perspective already know, namely, that I am a brain in a vat.

The conjunction of epistemic truth with internalist justification thus appears to make knowledge either too difficult (if *BV* and *RS* are both seen as non-ideal) or too easy (if they are both seen as ideal). In this latter case — which is the one we are now discussing — the attempt to rule out some wildly evidence-transcendent sceptical alternatives to (what we take to be) the actual world involves the dismissal of other evidence-transcendent alternatives which there is no reason to believe are either analytically or metaphysically impossible. Brain-in-a-vat situations like *BV* are merely a very special illustration of the fact that it is our current scientific conceptions that imply the possibility that many states of affairs might obtain unrecognizably to us. For our ability to decide the truth-value of many sentences about the world around us depends on many circumstances which are *naturally* contingent, in the sense that they are not entailed by anything that we could regard as laws of nature. With reference to sentences about the primary qualities of objects 'in the garden, in Antarctica or in the Andromeda galaxy', Loar (1987, 97) notes that our ability to decide their truth-value depends, among other things, on the arrangement of our neural pathways and on the fact that the regions of space through which the relevant light or sound must travel lack distorting properties. In the case of the sentence 'I am in the LSE library writing the second chapter of my Ph.D. thesis', my ability to decide its truth value (as opposed to merely being deluded about it) depends, among other things, on my not being the victim

of the experiments of evil super-psychologists. The possibility that I am such a victim is not the sort of hypothesis that can be excluded on purely a priori grounds (if it can be excluded at all). No epistemological theory excluding a priori this sort of hypothesis will be more plausible than a theory excluding a priori that the regions of space that separate us from objects in the garden, in Antarctica or in the Andromeda galaxy may have distorting properties. And this seems to me sufficient reason for rejecting the claim that *BV* and *RS* are both epistemically ideal situations.

Still, one could be tempted to escape the conclusion that epistemic truth is not going to be more accessible than correspondence truth by saying that the concept of an ideal epistemic situation refers to a situation which in principle can never be realized in the actual world (as Putnam does in his 1981, 55, but not in his 1988 and 1990). This would pre-empt my argument by granting its point: if 'we cannot really attain epistemically ideal conditions' (Putnam 1981, 55), then epistemic truth proves as inaccessible as correspondence truth allegedly is, and the possibility of 'sceptical worlds' is so much of a worry for the supporter of epistemic truth as it is for the supporter of correspondence truth. Adding that, although we cannot really attain an ideal epistemic situation, we can nevertheless approximate it to a very high degree of approximation, just as we can approximate a 'frictionless plane' to a very high degree of approximation (see Putnam 1981, 55; 1983, 84), will not do either. For if we do not know (and according to Putnam's own standards in the example under discussion we *do not know*) how to tell whether a real epistemic situation approximates the ideal one 'to a very high degree of approximation' or not, epistemic truth remains as inaccessible as correspondence truth allegedly is.

Saying that the concept of an ideal epistemic situation should not be taken as referring to the epistemic situation of an individual, but to the epistemic situation of a community (e.g. the scientific community, which is Peirce's, but not Putnam's idea), will make it possible to appeal to a 'larger' perspective, from which my own situation will indeed appear far from epistemically ideal. However, this fact will not refute the point that epistemic truth is quite incapable of making such simple truths as 'I am in the LSE library writing the second chapter of my Ph.D. thesis' epistemically accessible to individual cognitive subjects.

Therefore I think it is legitimate to conclude that the acceptance of a plausible degree of idealization is incompatible with the goal of devising a satisfactory epistemic

conception of truth, if that conception is to make truth epistemically more accessible (but not *too* accessible!) to human beings than correspondence truth allegedly is.

2.5 From Recognizability to Accessibility

As a matter of fact, Putnam has recently qualified his epistemic theory of truth by suggesting that not only truth depends on rational acceptability, but rational acceptability depends on truth: 'whether an epistemic situation is any good or not typically depends on whether many different statements are *true*' (Putnam 1988, 115). How are we to take this assertion? Is Putnam suggesting that the truth of those statements does not depend on their rational acceptability? If he were, then of course he would be giving up his former conception of truth and saying that rational acceptability depends on the non-epistemic truth of these statements. Surely this cannot be the case. So what is Putnam doing? I believe he is taking some cautious steps in the direction of an externalist theory of epistemic justification. The truth of the 'many different statements' which contribute to determine whether a given epistemic situation is ideal is of course *epistemic* truth. But a given epistemic situation may qualify as ideal even if the truth of those statements is not *known* to the cognitive subject.

This seems a most sensible move to do. The internalist philosopher is now in a position to make such simple truths as 'I am in the LSE library writing the second chapter of my Ph.D. thesis' epistemically accessible to individual cognitive subjects: everything she needs to do is to state that *externalist* epistemic justification (as described in the last section) is the only sort of epistemic justification that is necessary for knowledge. If the internalist philosopher is willing to state that, I may then be said to *know* that I am in the LSE library writing the second chapter of my Ph.D. thesis if I am, and the subjunctive fact that I would still believe that to be the case even if I were in fact a brain in a vat in the countryside will not be allowed to rob me of any knowledge that I actually have.

This sounds all very well and good. But does it mean that the argument given in the last section has finally been circumvented and that epistemic truth has been shown to be epistemically more accessible than correspondence truth? Not at all. The argument of the last section was not meant to establish that correspondence truth is epistemically

more accessible than epistemic truth or that endorsement of the latter will make truth epistemically inaccessible and knowledge impossible. I am happy to grant that one can develop an externalist view of epistemic justification on the basis of an epistemic notion of truth and that the resulting epistemology need not make epistemic truth any more inaccessible than correspondence truth. What I claim is simply that, by taking this line, the upholder of epistemic truth will not gain any epistemological benefit which is not also available to the upholder of correspondence truth.

This is because, *ceteris paribus*, epistemic truth can prove more accessible than correspondence truth only if it is defined in terms of an *internalist* notion of epistemic justification. The whole point of defining truth in epistemic terms is that the provability, verifiability, rational acceptability, or warranted assertability of a given sentence is typically taken to be a *recognizable* property, whereas the sentence's correspondence to a knowledge-independent reality is argued to be an *unrecognizable* property. This is the only reason why epistemic truth may be thought to be intrinsically more accessible than correspondence truth. And if truth were pure and simple verifiability, this would indeed be the case. But as soon as reference to an ideal epistemic situation is built into the definition of epistemic truth, the provability, verifiability, rational acceptability or warranted assertability of a given sentence is bound to become as unrecognizable a property as the sentence's correspondence to knowledge-independent reality.

We have seen that one can then bite the bullet and say that an ideal epistemic situation need not be *known* to be ideal to qualify as such. If I happen to be in an ideal epistemic situation for the purpose of knowing that p, then it may very well be the case that I do know that p, or, in other words, that I have epistemic access to p's (epistemic and yet unrecognizable) truth. Perhaps these considerations fall short of a 'proof' that I know that p, but surely they will bring one some epistemological comfort. And what else might such limited creatures as we are realistically hope for?

For all the apparent plausibility of this sort of position, one should be very clear about what it does and what it does not provide. First, it does not provide the refutation of the sceptic that Putnam's 'Brain in a vat' argument was supposed to provide. Second, it does not provide a *recognizable*, but only an *accessible* concept of truth (and of course the access to truth that interests us cannot be a matter of sheer chance, but has to be an *epistemic* matter). Third, and most important, it does not provide an account of the accessibility of truth which is not available to the upholder of correspondence truth. For

as soon as the accessibility of truth comes to be regarded as compatible with its recognition-transcendence, it becomes clear that the upholder of epistemic truth cannot be any better off than the upholder of correspondence truth. There is no reason why the latter, unlike the former, should not be allowed to avail herself of an externalist view of epistemic justification by saying, for example, that S has epistemic access to the truth that *p* just in case her (correspondence-true) belief that *p* is reliably produced by the fact that *p*. And this will enable the upholder of correspondence truth to reap the very same epistemological benefits the anti-realist advertises her own system as providing without having to substitute some sort of Kantian phenomenal world for the knowledge-independent reality we usually take our knowledge to be about.

I think that epistemology can do without a recognizable notion of truth and a rigorous refutation of the sceptic, as long as it can make sense of the idea that our access to truth may be epistemic and not merely serendipitous. But I am not convinced that it should give up the idea that if there is anything that we know, that is reality as it is 'in itself'. In the next chapters I shall then explore in greater detail the way in which an externalist view of epistemic justification and/or knowledge can support an argument for the *epistemic* accessibility of correspondence truth and knowledge-independent reality.

Knowledge Without Certainty

Nowadays nobody believes any longer that our knowledge of the world may be rendered absolutely certain and safe from revision. And Putnam is right to remind us that no God's eye point of view is accessible to human beings. This means, I think, that if one wants to argue that correspondence truth and knowledge-independent reality are in fact epistemically considerably more accessible than upholders of epistemic truth are wont to grant, one will have to give up the 'internalist' conception of knowledge, i.e., the conception of knowledge that is commonly ascribed to modern (post-Cartesian) epistemology.

Contemporary philosophical jargon provides no less than three different senses in which 'internalism' can be contrasted with 'externalism'. First of all, one must be careful not to confuse the internalist conception of knowledge with the 'internalist' perspective advocated by Putnam over and against the 'externalist' perspective of a God's eye view of reality. We saw in the last chapter that the point of Putnam's advocacy of the internalist perspective was in fact the rejection of the conjunction of metaphysical realism, correspondence truth, and the thesis that there is exactly one true and complete description of the way the world is. For the sake of clarity, I shall call Putnam's version of the internalism/externalism dichotomy the *metaphysical* version.

A second version of the internalism/externalism dichotomy is also due to Putnam (see Putnam 1975, 223-227), although in this case he cannot be held responsible for the terminology (nor can the other father of the doctrine, Tyler Burge, in his 1979). This second version belongs to the *philosophy of mind*. The issue is, roughly, what determines

the nature of intentional states (or, if you like, propositional attitudes). Internalist theories of mind claim that the nature of intentional states is entirely determined by factors which are 'internal' to the subject. Externalist theories claim that important aspects of the nature of intentional states (e.g., their referents) are affected by such 'external' factors as the nature of the subject's environment. Putnam's Twin Earth argument was originally put forward as a proof of the view that '"meanings" just ain't in the *head*' (Putnam 1975, 227). So Putnam, while an internalist on the issue of realism, is to be counted as an externalist in the philosophy of mind. However, we need not spell out this issue in detail, because what is relevant to our topic is rather the third version of the dichotomy, to which we now turn our attention.

The *epistemological* version of the internalism/externalism dichotomy, which has been the focus of extensive philosophical debate in the last two decades or so, emerged in the context of the analysis of the concepts of knowledge and epistemic justification. In section 2.4 we saw that the thrust of an internalist view of epistemic justification is that a person cannot be justifed in believing that p if she has no 'cognitive grasp' of whatever it is that justifies her believing that p. In the same way, we can say that the thrust of an internalist view of knowledge is that a person cannot know that p if she has no 'cognitive grasp' of whatever it is that turns her believing that p into her knowing that p. This is the sense of 'internalism' on the basis of which it has become commonplace to describe post-Cartesian epistemology as 'internalist', and it is with this kind of internalism that I will be mostly concerned in this chapter.

My detailed definitions of internalism and externalism in the theories of knowledge and epistemic justification are presented in section 3.2. They are embedded in a sort of 'historical' sketch which should help to put epistemological internalism in perspective and to explain its appeal on many past and present philosophers; in particular, section 3.1 addresses the significance of the epistemological dogmatism of the founder of modern epistemology, René Descartes, and the German phenomenologist, Edmund Husserl, whereas section 3.3 explains how epistemologists in the Cartesian tradition may come to regard an internalist conception of knowledge as the most natural way to spell out the plausible idea that in many circumstances S's knowing that p must involve S's ability to tell the difference between her knowing that p and her merely apparently knowing that p. An internalist view of knowledge is, however, exposed to serious difficulties, and section 3.4 contains a discussion of what is often taken to be the

major theoretical obstacle on the way of an internalist view of knowing, namely, the 'vertical' epistemic regress problem.

3.1 Realism From a God's Eye Point of View

Even if the popularity of epistemological externalism has recently increased, epistemological internalism is commonly regarded as the 'orthodoxy' of modern epistemology. In order to understand the point of the commitment to epistemological internalism of most modern epistemologists, it may be useful to start from what might be called - in Dewey's phrase — the 'quest for certainty'.

Certainty can be sought for its own sake or for the sake of truth. People with a dogmatic temperament find a psychological comfort in the possession of certainty (Peirce went so far as to see in the 'fixation of belief' one of the natural goals of human beings, supposedly seeking relief from the uncomfortable state of doubt), but *mere* certainty is not necessarily connected with truth. 'Certainty' may no doubt mean different things. One should distinguish the 'objective' certainty of analytical (logically self-guaranteeing) truths from the 'subjective' certainty attached to those beliefs which are experienced as indubitable by their possessor (see Haack 1979a, 53; 1979b, 327). Objective certainty displays an obvious connection with truth, whereas subjective certainty may well be a psychological (or sociological) artifact devoid of any cognitive value.

While many nineteenth century logicians may be suspected of having identified psychological compulsion with logical truth, thus obscuring the distinction between certainty and truth, the father of modern epistemology, René Descartes, was altogether clear that the primary aim of his search for certainty was the discovery of something quite different from certainty itself: the aim of his search was the discovery of substantial *truths* about the world. The point of what Bernard Williams calls Descartes' 'project of pure enquiry' was to put aside every practical concern which could constrain the cognitive enterprise in order to devise an effective method for acquiring true, and only true, beliefs about the world. But an effective method for acquiring true, and only true, beliefs about the world is a method the correct application of which *guarantees* truth. And this 'comes to the requirement that the beliefs which the method generates should be *certain*' (Williams 1978, 48 f.). This explains why Descartes' search for truth

turned into a search for certainty, even though his primary concern was with the former rather than the latter (Williams 1978, 200). A famous passage from Descartes' *Meditations on First Philosophy* reads:

> Anything which admits of the slightest doubt I will set aside just as if I had found it to be wholly false; and I will proceed in this way until I recognize something certain, or, if nothing else, until I at least recognize for certain that there is no certainty.
>
> (Descartes 1984-85, II, 16)

Descartes eventually thought to have found what he was looking for in the proposition, *I am, I exist*[1], which he took to be self-evident and then used as the uncontrovertible foundation of his epistemological construction. In order to carry out his project, he had to appeal to a bridge-principle saying that whatever is perceived very clearly and distinctly — i.e. with the marks of certainty — must be *true* (Descartes 1984-85, II, 24). It should be emphasized that Descartes did not derive the content of his knowledge of the external world from a limited set of self-evident truths related to the *cogito*. Rather, he used those truths to 'rehabilitate' the deliverances of the senses and to reject sceptical doubts. This he could do only by appealing to the existence and benevolence of God: the vindication of the bridge-principle connecting certainty with truth rests (circularly?) on the impossibility that a benevolent God should ever deceive those who make a careful use of their power of reasoning (see Descartes 1984-85, II, 11 and 37; also the passage from Descartes' *Principles of Philosophy* quoted in section 1.2).

Descartes can be fairly regarded as a paradigmatic example of that kind of epistemological strategy, dominant throughout the history of modern philosophy, that regards certainty as an essential feature of knowledge. The link between certainty and truth is established, in Descartes' philosophy, by God's role in vouching for the reliability of our cognitive faculties. But when the appeal to God no longer appeared as a viable epistemological option, it became necessary to provide new foundations for the link between certainty and truth. An interesting strategy for grounding this link without

---

[1] '*I am, I exist*' is the proposition Descartes claims to be necessarily true ('whenever it is put forward by me or conceived by my mind') in the Second Meditation. This is of course the *cogito* argument, but in the *Replies* Descartes warns the reader against an easy misconstruction of his formula, *cogito, ergo sum*: 'When someone says "I am thinking, therefore I am, or I exist", he does not deduce existence from thought by means of a syllogism, but recognizes it as something self-evident by a simple intuition of the mind' (Descartes 1984-85, II 100).

relying on the existence and benevolence of God can be found in Husserl's *Logical Investigations*.

Husserl lists four different meanings of 'truth', but he regards as basic the concept of truth as 'the *adaequatio rei ac intellectus*' (Husserl 1970, II, 670). According to Husserl, truth makes itself present in that kind of objectifying act in which 'the object is not merely meant, but in the strictest sense *given*, and given as it is meant' (Husserl 1970, II, 765). This quotation reflects Husserl's own distinctive terminology. Note first of all that 'object' is used in a very broad sense, including states of affairs as well as individuals. Husserl conceives truth as the objective correlate of a psychological act in which a meaning-intention comes to perceptual fulfilment. Note that he does not speak of sentences, but of meaning-intentions. Meaning-intentions are temporally located psychological acts, but Husserl is well aware of the dangers of psychologism, and describes their contents, in a Fregean vein, as 'the self-identical meaning that the hearer can grasp even if he is not a percipient' (Husserl 1970, I, 290). When the cognitive subject is not facing the intended object, her meaning-intention is a *mere* meaning-intention, that is, void of intuition. But when the intended object comes before the subject, her meaning-intention may be fulfilled, if the object is *given* to her by perception (which Husserl understands in a suitably liberalized fashion) in exactly the same way as it is meant. If so much happens, Husserl says that an 'identifying act' takes place, which has as its object 'the *full agreement* of what is meant with what is *given as such*', that is, '*being in the sense of truth*, or simply *truth*' (Husserl 1970, II, 765).

What is particularly relevant to our enquiry is Husserl's view of the relation between truth and self-evidence:

> This agreement [i.e., truth] we *experience* in self-evidence, in so far as self-evidence means the actual carrying out of an adequate identification.
>
> (Husserl 1970, II, 765)

According to Husserl, whenever we run up against this agreement we also have the possibility of laying it before our consciousness:

> Truth is indeed 'present'. Here we have always the *a priori* possibility of looking towards this agreement, and of laying it before our intentional consciousness in an adequate percept.
>
> (Husserl 1970, II, 766)

When a meaning-intention is given the fullness of the object itself, 'the *adaequatio rei ac intellectus* [...] is itself given, to be directly seized and gazed upon' (Husserl 1970, II, 670).

Husserl's characterization of 'self-evidence' can be seen as an attempt to provide a non-psychologistic explication of the notion of 'certainty'. It is clear that his concept of 'self-evidence' precludes any interpretation of certainty as a mere 'feeling' contingently attached to the act of judgement (see Husserl 1970, II, 769). Although certainty *is* an experience of the knowing subject, it is not incidental to the presence of truth. The relation between truth and certainty becomes so intimate that Husserl is eventually led to identify knowledge with certainty:

> The synthesis of fulfilment achieved in this limiting case [the case of an objectively complete adequacy of the meaning-intention to the object itself] is *self-evidence or knowledge in the pregnant sense of the word.*
>
> (Husserl 1970, II, 670)

For Husserl certainty is not merely desirable as a sufficient condition for truth, nor is it simply an essential feature of knowledge: certainty is knowledge itself, *in the pregnant sense of the word.* Truth cannot be known without the *stigmata* of certainty.

What I am most interested in emphasizing, though, is the fact that Husserl appears to believe that acceptance of correspondence truth implies commitment to the accessibility of an epistemic standpoint which can be aptly described, in Putnam's phrase, as a 'God's eye point of view'. Of course he does not use these words, but what else could he mean when he says that in the synthesis of fulfilment 'the object is not merely meant, but in the strictest sense *given*, and given as it is meant'? To be sure, Husserl allows for increasing degrees of intuitive fulfilment. Yet he contends that what is given in the experience of fulfilment it not a mere phenomenon, but the *Ding-an-Sich*: the experience of fulfilment 'is represented by the words: "This is the thing *itself*"' (Husserl 1970, II, 720). This means that when Husserl writes that in the synthesis of fulfilment the object is 'given as it is meant', he does not want to imply that it is not given as it is in itself. His theory of intentionality (which is reminiscent of, but not identical with, Brentano's theory) enables him to say that what is *given* in perception is not merely a representation or an idea, but the object itself (see Husserl 1970, II, 552-596, especially the section 'Critique of the "image-theory" and the doctrine of the

"immanent" objects of acts'). Husserl's contention is then that in the synthesis of fulfilment there is a full agreement between the object as it is in itself and the object as it is meant. If truth is *adaequatio rei ac intellectus* — so goes Husserl's assumption — knowledge at its best must be the self-evidence of that *adaequatio*. Husserl's contention is that knowledge in this sense is indeed possible, because in the synthesis of fulfilment 'this *adaequatio* is itself given, to be directly seized and gazed upon' by the cognitive subject. This is tantamount to saying that the cognitive subject has epistemic access, in certain circumstances, both to her description of the world and to the world as it is in itself. Moreover, Husserl tells us that she can compare their features, and also that, when the object is given as it is meant, she can perceive this agreement in the experience of certainty.

Now, what moral can we draw from this brief discussion of Descartes' and Husserl's epistemologies? We have seen how Descartes sought to vindicate the power of (carefully used) human reason by appealing to the benevolence and trustworthiness of God. The fact that human reason could attain genuine knowledge of reality was depicted by him as a gracious gift of the divinity. But when the appeal to God no longer appeared as a viable epistemological option, it began to look as if (metaphysical) realism could accommodate the possibility of human knowledge only if human beings themselves could be attributed the possession of that God's eye view of reality (or at least of some aspects of it) that Descartes had thought they could borrow from God. Unlike Descartes, in the *Logical Investigations* Husserl could not take refuge in the existence of a benevolent God to ensure the correspondence of human beliefs (meaning-intentions) with a knowledge-independent world: his project of a rigorous foundation of scientific knowledge led him to endow the cognitive subject herself with a direct access to the objects of her enquiry and with the capacity of *seeing* whether her beliefs match the way the world is.

We may regard Husserl's view, with its commitment to the epistemic accessibility of a God's eye view of reality, as a striking illustration of precisely that kind of philosophy that Putnam strives to reject in his attack against metaphysical externalism. On the other hand, Husserl's view also provides a striking illustration of that kind of internalist epistemology that I believe a metaphysical realist need not be committed to. One of the main points of this work will be that commitment to the accessibility of a God's eye view of reality should not be seen as part and parcel of (non-sceptical)

realism. And I shall suggest that the fact that this is not always recognized is a consequence of the age-old allegiance of traditional epistemology to the quest for certainty and/or to the 'internalist' perspective which has outlived that quest when philosophers began to realize the fallibility of human knowledge. But before I embark on a discussion of epistemological internalism, I have to say a few more things about the general characters of dogmatic epistemologies.

Both Descartes and Husserl take epistemic accessibility to involve certain, or self-evident, knowledge of the object of inquiry. Descartes insists that true knowledge, as opposed to mere persuasion, cannot be rendered doubtful, whereas Husserl describes a particular objectifying act (that is, self-evidence) in which the fulfilment of the necessary conditions for knowledge is experienced by the cognitive subject. Both Descartes and Husserl are therefore committed to the claim that whenever S knows that *p*, she can also know for certain that the conditions which must be satisfied for her to know that *p* are actually met. This is a general feature of dogmatic epistemologies, that is to say, of those epistemologies that assert (i) that human beings have access to (recognizably) infallible cognitive methods, and (ii) that the certainty guaranteed by such methods is an essential feature of knowledge.

Dogmatic epistemologies are apparently committed to the thesis that knowing that *p* includes knowing that one knows that *p*:

> We must recognize that whenever we know something we either do, or at least can, by reflecting, directly know that we are knowing it.
>
> (Prichard 1950, 86)

Following Hintikka, Chisholm (1989, 99 f.) refers to this thesis as to 'the KK principle'.

As it stands, the KK principle is unlikely to be true, because it seems clear that any plausible (internalist or externalist) account of knowledge will have to make room for the possibility that a person may know that *p* and yet lack the very concept of knowledge. Such a person will not understand the meaning of propositions like 'S knows that *p*', or 'S knows that S knows that *p*'. Assuming, reasonably enough, that a proposition cannot be known by a person unless it is *understood* by that person, someone lacking the concept of knowledge will be unable to know that she knows that *p* (Danto 1967). This is why Chisholm proposes to replace the KK principle with the 'objectivity principle':

> The objectivity principle tells us that, if a person knows a given proposition to be true, and if he also *believes* that he knows that propositions to be true, then he *knows that he knows* that proposition to be true.
>
> (Chisholm 1989, 100)

This substitution seems sensible enough. However, the objectivity principle is, as its predecessor, a conditional: 'If S knows that *p* and... , then...'. Cases of merely apparent knowledge which are evidentially indistinguishable from cases of actual knowledge are perfectly consistent with the truth of the objectivity principle: the fact that in such cases S can wrongly believe that she knows that *p* does not count as a counterexample to the principle. Therefore it seems clear that dogmatic epistemologists must be committed to a stronger principle. If true knowledge is to resist every attempt to render it doubtful, dogmatic epistemologists must deny that (true) certainty can ever be misplaced. Not only must people be able to know that they know that *p* when they do: they must be able to know that they don't know that *p* when they don't! I shall call this the 'infallibility principle':

> *Infallibility Principle*: For any *p*, if S understands the concept of 'knowledge' and reflects upon her own epistemic situation with respect to *p*, then:
> (a) if S knows that *p*, S cannot fail to know that she knows that *p*; and
> (b) if S does not know that *p*, S cannot fail to know that she does not know that *p*.

Nothing short of this principle can express the totality of the commitments of those dogmatic epistemologists who, like Descartes and Husserl, regard certainty as an essential — and attainable — feature of knowledge. For if S can, even after reflection, mistake cases of merely apparent knowledge for cases of actual knowledge, and if this is a wholly general predicament, then it is clear that every putative case of knowledge will be, for S, open to doubt.

The requirements the infallibility principle imposes on knowledge are very stringent; indeed, they are so stringent that nowadays it has become commonplace to reject them as unfulfillable. In section 3.3 I shall argue that, since the internalist conception of knowledge can be seen as a watering down of the infallibility principle, it is small wonder that after the giving up of the quest of certainty philosophers in the Cartesian tradition came to adopt such a conception of the nature of human knowledge as the most natural surrogate of a dogmatic epistemology. But before doing this, I have

to provide a more precise characterization of what is involved in epistemological internalism and externalism.

3.2 Internalism and Externalism in Epistemology

The definitions of epistemological internalism and externalism one can find in the relevant literature (see especially Alvin Goldman 1980, Bonjour 1980, Dancy 1985, and Alston 1989) are far from univocal. And yet I think there is something more than a mere family-resemblance among them. I shall presently introduce my own definitions; then I will briefly discuss their features and present a couple of examples that will help the reader to understand how they work.

I propose to say that an analysis of 'S knows that *p*' is 'internalist' ['externalist'] when it requires [when it does not require] that the satisfaction of the conditions that must be met for S to know that *p* be 'internal' to S's awareness, that is, when it requires [when it does not require] that S have epistemic access to the satisfaction of those conditions. Similarly, I propose to say that an analysis of 'S is justified in believing that *p*' is 'internalist' ['externalist'] when it requires [when it does not require] that whatever it is that makes S's belief justified be 'internal' to S's awareness, that is, that it be epistemically accessible to S.

The reference to what is 'internal' or 'external' to S's awareness helps to explain the labels 'internalism' and 'externalism', but it is by focusing on the concept of 'epistemic accessibility' that my characterization of epistemological internalism [externalism] can be best specified. Let us concentrate on epistemological internalism (epistemological externalism is just its negation). Two issues immediately present themselves. First of all, one needs an explanation of *how* epistemically accessible the relevant conditions must be in order to justify a claim to knowledge (or to epistemic justification). Does S need to *know* that those conditions are satisfied? Or does she merely need to have a *justified* belief that they are? Second, must S's knowledge (or justified belief) that those conditions are satisfied be actual or can it be merely potential? Different kinds of internalist epistemologies will issue from different answers to these questions.

On the actuality/potentiality issue, I think we can conveniently agree to regard the epistemic access required by internalism as merely potential, which seems the most sensible option to take if we do not want to exclude too many items from our inventory of knowledge and/or epistemic justification (people don't usually spend their time knowing or justifiedly believing that they know or justifiedly believe this or that). Accordingly, if John knows [justifiedly believes] that his sister is in Berlin, we shall not construe an internalist analysis of his knowledge [justified belief] as requiring that John actually know or justifiedly believe that the relevant conditions for him to know [justifiedly believe] that his sister is in Berlin are satisfied. We shall be content to impose the requirement that he *can* know or justifiedly believe that the conditions which are to be met if he is to know [justifiedly believe] that his sister is in Berlin are actually satisfied. This means, for example, that if somebody asks John whether he really knows [justifiedly believes] that his sister is in Berlin (and John understands the question), John's knowing [justifiedly believing] that his sister is in Berlin must enable him to answer that he does, and the belief expressed in his answer must qualify as knowledge [be epistemically justified][2].

As for the question of the *degree* of epistemic access — knowledge or justified belief? — required by epistemological internalism, any answer would now be premature. Different solutions to this question yield different versions of internalism, none of which is obviously implausible. In this section I shall confine myself to an illustration of the features of various possible solutions. This illustration will involve the consideration of two popular accounts of knowledge. I shall ignore accounts of epistemic justification: since the concept of epistemic justification is usually employed in the analysis of the concept of knowledge, one is unlikely to come across a construal of the internalist requirement on epistemic justification having 'knowledge' as the *explanans* of 'epistemic accessibility' (however I might as well anticipate that, failing to take epistemic justification as a necessary condition of knowledge, this is precisely the option I myself happen to favour).

First take the traditional (and largely discredited) analysis of knowledge as justified true belief:

---

[2] Here I presuppose that belief is necessary for knowledge, which of course has been questioned, e.g., in Austin 1946 and Radford 1966. However, it is clear that the requirement that S have epistemic access to the satisfaction of the conditions that are to be met for her to know [justifiedly believe] that *p* does not involve, in itself, this presupposition.

S knows that *p* **iff** 1) *p* is true
2) S believes that *p*
3) S is justified in believing that *p*

Is this analysis internalist or externalist? That depends (1) on our concept of justification and (2) on our concept of epistemic accessibility. If our concept of 'epistemic accessibility' is not too demanding (so that satisfaction of condition 3 may be seen as providing epistemic access to the satisfaction of condition 1), and if our concept of justification is itself internalist (so that S may be thought to have epistemic access to the satisfaction of 3), this analysis can be described as 'internalist' (on the further, reasonable assumption that S has epistemic access to the satisfaction of condition 2). On the contrary, if (a) our concept of epistemic accessibility is not satisfied by mere epistemic justification, but requires knowledge, or if (b) our concept of justification is externalist, the JTB account of knowledge will come out as externalist. If our concept of epistemic accessibility requires knowledge, satisfaction of condition 3 will not provide epistemic access to the satisfaction of condition 1 (unless we commit ourselves to the implausible assumption that one cannot justifiedly believe a false proposition). If our concept of justification is externalist, S will not have epistemic access to the satisfaction of condition 3.

The second account of knowledge I want to examine is provided by the epistemological position which goes under the name of 'reliabilism':

S knows that *p* **iff** 1) *p* is true
2) S believes that *p*
3) S came to believe that *p* by means of a reliable belief-forming process

A belief-forming process is completely reliable when it produces only true beliefs. According to Armstrong (1973, 166), for example, 'there must be a *law-like connection* between the state of affairs *Bap* and the state of affairs which makes "*p*" true, such that, given *Bap*, it must be the case that *p*'. But the reliability of a belief-forming process can be defined in many different ways; for example, one can regard as reliable also those processes which produce true beliefs more often than not.

Even if most reliabilist analyses of knowledge are advanced by their proponents as externalist, nothing in the above definition excludes the possibility of an internalist

interpretation. It is clear that condition 3 may be satisfied without S having any epistemic access to its satisfaction. In fact, S will be mostly unaware of the particular belief-forming process that led her to believe that *p*, let alone of its being reliable or not. Yet, there seem to be no *prima facie* contradiction in requiring that S have the possibility of knowing or justifiedly believe that the process by which she came to believe that *p* is indeed a reliable belief-forming process. If such a requirement is imposed, and if the reliability of that belief-forming process is seen as granting epistemic access to the satisfaction of condition 1, this analysis of knowing may well come out as internalist. However, I do not know of any proponent of a reliabilist analysis who takes a straightforwardly internalist standpoint. People subscribing to reliabilist analyses of knowing are very careful not to require that the cognitive subject have epistemic access to the reliability of her belief-forming processes. The point of their proposing a reliabilist analysis is precisely to emphasize that no such access represents a necessary condition for knowledge. All reliabilist analyses of knowing that I am familiar with are thus to be counted as externalist[3].

For the sake of brevity, I will sometimes speak of 'transparent' and 'opaque' instances of knowledge. I will say that S's knowing that *p* is epistemically 'transparent' ['opaque'] to S [R] just in case S [R] has [has not] epistemic access to the satisfaction of the conditions that must be met in order for S to know that *p* (that is, just in case S [R] can [cannot] know/justifiedly believe that those conditions are satisfied)[4].

I have introduced 'R' alongside with 'S' because there may be circumstances in which S's knowing that *p* is epistemically opaque to S but transparent to some other individual R, or transparent to S but opaque to R. It is an immediate consequence of this definition that every internalist analysis of knowing will require of any true instance of knowledge that it be 'transparent' to the cognitive subject to whom (or to which) it

---

[3] Somewhere between the JTB and reliabilist accounts of knowing one can find analyses of 'S knows that *p*' which attempt to handle Gettier's (1963) examples of justified true beliefs which do not qualify as knowledge by complementing, rather then replacing, the justification condition. These include the so-called 'indefeasibility' accounts, according to which knowing is, roughly, having a justified true belief the epistemic justification of which cannot be 'defeated' by any true statement the truth of which the knowing subject is currently unaware of (see, e.g., Lehrer 1990, esp. ch. 7). According to my definition of epistemological internalism [externalism], such accounts come out as externalist, because they do not require the knowing subject to have epistemic access to the fact that her justification is undefeated (indeed, it is a consequence of those accounts that the knowing subject will typically lack epistemic access to the fact that her justification is undefeated).

[4] I will not regard as a necessary condition for S's knowing that *p* to be epistemically transparent to S [R] that S [R] have epistemic access to the fact that those conditions *are* the conditions for S to know that *p*. S's knowledge that *p* can thus be transparent to S [R] even if S [R] is not an epistemologist!

belongs. On the other hand, an externalist analysis will not necessarily exclude the existence of instances of knowledge which are transparent to the cognitive subject to whom (which) they belong, but it will also allow for the possibility of genuine instances of knowledge which are 'opaque' to the cognitive subject to whom (which) they belong.

3.3 Epistemological Internalism and Knowing That One Knows

Having clarified what I mean by 'internalism' and 'externalism' in the theories of knowledge and epistemic justification, I can now present the awaited explanation of the fact that philosophers in the Cartesian tradition have mostly retained an internalist conception of knowledge even after the decline of epistemological dogmatism. But I have to warn the reader that in this section I will tell only the first half of the story; the second half — the *good* reasons for espousing an internalist theory not of knowledge, but of epistemic justification — will emerge only in chapter 5.

That epistemological dogmatism involves an internalist conception of knowledge is an immediate consequence of the definition of knowledge-internalism. One might then feel inclined to believe that the bankruptcy of epistemological dogmatism justifies a wholesale dismissal of all knowledge of one's knowledge. The fact that certainty is now commonly regarded as an unattainable goal, and hence as an undesirably strong requirement to be set upon (empirical) knowledge, may seem to facilitate the task of those philosophers who conceive truth as correspondence to a knowledge-independent reality: if knowledge does not require certainty, the fact that we can never make certain that a given sentence is correspondence-true will not rule out the possibility that we may *know* that the world is as that sentence describes it.

Williams (1978, 45), for example, has a reliabilist account of knowing which is definitely externalist and according to which S knows that *p* if S truly believes that *p* and her belief has been '*appropriately* produced in a way such that beliefs produced in that way are generally true'[5]. This account enables him to say that there is 'no obvious impossibility in the idea that the natural sciences should be able to give absolute explanations of a determinate and realistically conceived world' (Williams 1978, 302).

[5] Williams presents his account as providing merely sufficient conditions for knowledge, because he does not want to go into the question whether everyone who knows has to have a belief.

But having exposed as hopeless Descartes' search for certainty, Williams goes on to argue that our epistemic objective is just to gain as much 'absolute knowledge' (that is, knowledge of 'what is there *anyway*') about the world as we can, and that to 'ask not just that we should know, but that we should know that we know, is [...] to ask for more — very probably for too much' (Williams 1978, 303). If I understand him correctly, the price he is willing to pay for being entitled to the claim that the world accessible to our knowledge need not be (as he puts it) relative to our language and conceptual scheme is not only the giving up of certainty, but of all knowledge of our knowledge. Isn't this too much?

One could deny that this is too much by arguing that epistemologists are the only sort of people who could find it appropriate to make second order knowledge claims like, 'I know that I know that *p*'. If this is so, then it seems clear that for all practical purposes we would not be any worse off for the lack of knowledge of our knowledge. However, epistemologists in the Cartesian tradition are unlikely to be impressed by this sort of argument. They will retort that it is hard to conceive that S might possibly know that *p* without having at least *some* degree of epistemic access to her own epistemic situation. For how could one know that *p* if one had no reason at all to believe that one did? How could one know that *p* and at the same time be totally incapable of telling the difference between knowing that *p* and merely *apparently* knowing that *p*?

This sort of thoughts are likely to have led several epistemologists to view certainty as a sort of regulative ideal which, although unattainable, can nevertheless be approximated to some extent by human knowledge, in so far as human cognitive subjects maintain some degree of epistemic access to their own epistemic situation. Giving up certainty as a necessary condition for knowledge involves giving up the infallibility principle. This follows from the fact that if epistemic justification cannot guarantee truth, then S may be justified in believing the false proposition that *p*. If so, then of course S will not know that *p*, but even if she understands the concept of 'knowledge' and reflects on her own epistemic situation with respect to *p*, she may very well fail to know that she doesn't, for if her justification is strong enough, she will most likely believe that she does. Acceptance of the infallibility principle is thus incompatible with acceptance of a fallibilist (i.e., non-dogmatic) epistemology[6]. However, giving up certainty as a

[6] This result does not presuppose endorsement of the JTB account of knowing; for the infallibility principle is incompatible with the mere claim that one can, after reflection, wrongly believe that one knows

necessary condition for knowledge does not necessarily involve giving up the objectivity principle as well. But the conception of knowledge one is likely to end up with by giving up the infallibility principle while retaining the objectivity principle is precisely an *internalist* watering down of the dogmatic conception of knowledge. And this, I surmise, is the motivation why several philosophers have given up the quest for certainty but retained some kind of internalist epistemology according to which, if we are to have any knowledge at all, we need to have at least some *probable* grasp of our own epistemic situation. For commitment to an internalist theory of knowledge is a direct consequence of commitment to the objectivity principle (if the required epistemic access to the satisfaction of the conditions that must be met for S to know that *p* is spelled out in terms of knowledge, the reverse is also true. If it is spelled out in terms of epistemic justification, then S's knowing that *p* — and believing that she does — can fail to involve S's knowing that she knows that *p* only if one supposes that S's being justified in believing that the conditions that must be met for her to know that *p* are satisfied — as indeed they are — can fail to qualify as knowledge of her knowing that *p*. This supposition is not unreasonable, but one has to assume that there is some condition that is necessary for knowledge that S's being justified in truly believing that the conditions that must be met for her to know that *p* are satisfied does not itself satisfy. And this assumption is typically resisted by epistemologists in the Cartesian tradition).

Contemporary commitment to an internalist analysis of knowledge can thus be seen as a way to preserve the intent of traditional epistemology (which took the quest for certainty as an essential feature of every cognitive effort), while at the same time relaxing its criteria of epistemic accessibility.

Now, I must say that I do share the idea that the bankruptcy of epistemological dogmatism cannot justify a wholesale dismissal of all knowledge of one's knowledge, but my reasons for this claim are a bit different from the reasons I have been describing so far. My reasons have to do with the fact that people do make second order knowledge claims outside epistemological circles, because second order knowledge claims are in fact an essential ingredient of the activities by which we seek to improve our epistemic situation and gain some control over our inquiries.

Consider the following conversation (which was suggested to me by Nancy Cartwright): 'I am six feet tall'. 'How do you know?'. 'My daughter measured me. So

---

that *p*.

I do know'. 'But she is extremely unreliable'. 'Not in this case. I watched her and she was very careful, so *I do know that I know*'. There are many such cases in which we describe ourselves as knowing that we know this or that. These cases typically involve a commitment to the claim that the process or method that originated the ground-level belief is cognitively reliable. Of course one can always imagine sceptical circumstances in which these 'second order' knowledge claims might be mistaken. But this is a completely general fact that applies to all sorts of (empirical) knowledge claims irrespective of their being first or second order. Therefore I think that here we have a reason why an adequate analysis of the concept of knowledge should account for second order knowledge claims rather than simply dismiss them as blatant mistakes or epistemological abstrusities. This is because an analysis of knowledge ruling out all knowledge of one's knowledge could hardly account for the fact that we can often improve our epistemic situation by assessing the reliability of our cognitive methods and strategies. Not only do I know that I weigh nine stones four: I know which methods I can trust to give me knowledge of my weight (this is why I know that I know that I weigh nine stones four if I have just weighed myself on my doctor's scales, while I don't if I have just weighed myself on my own bathroom scales). The fact that in many cases we can evaluate the reliability of our cognitive processes and methods (and revise our beliefs accordingly) is the reason why I believe that a satisfactory epistemology cannot do without knowledge of one's knowledge.

But is an internalist account of knowing the only way of rendering the fact that in certain circumstances not only can human beings know that p, but also tell the difference between their knowing that p and merely apparently knowing that p? Must the existence of second, and maybe higher, order knowledge claims necessarily lead one to require that *all* genuine instances of empirical knowledge be epistemically transparent to the subject to which they belong? I do not think this is the case. I believe that it is possible to devise an externalist account of knowing according to which (i) human beings can often turn out to know that they know that p, and yet their (capability of) knowing that they know that p (ii) does not represent a necessary condition of their knowing that p, nor (iii) does it enable them to establish p's truth beyond any possible doubt. I take it that the capacity of satisfying condition (i) is crucially important. For even if an externalist analysis of knowledge failing to satisfy requirement (i) might be thought to suffice to make correspondence-truth not only accidentally, but *epistemically* accessible

to our efforts, it seems to me that its failure to allow for second order knowledge claims would prevent it from making sense of the fact that we can gain some control over our inquiries by assessing the reliability of our cognitive methods and strategies.

In the next chapter I shall then discuss two externalist accounts of knowing which meet (one of them after a minor revision) this requirement. Meanwhile, I shall present an argument that is purported to show that internalist accounts of knowing are bound to face unsurmountable difficulties if they are not complemented by an *externalist* account of epistemic justification.

3.4 The Vertical Epistemic Regress Problem

The main predicament of internalist accounts of knowing is represented by what I shall call the 'vertical' epistemic regress problem, to distinguish it from the more commonly discussed 'horizontal' epistemic regress problem.

The horizontal epistemic regress problem stems from the requirement that our beliefs be epistemically justified. Anyone will agree that beliefs can be justified inferentially, i.e., by deriving them from other beliefs; but inferential justification cannot be carried on *ad infinitum*. So the horizontal epistemic regress problem is usually developed as an argument in support of some kind of foundationalist epistemology. In other words, it is maintained that there must be some beliefs which are justified non-inferentially if there are to be inferentially justified beliefs at all. On the other hand, the vertical epistemic regress problem stems from the internalist requirement that the conditions that must be satisfied for S to know that *p* be epistemically accessible to S.

Alvin Goldman (1980, 38-41) has a painstaking discussion of the vertical epistemic regress problem, but his argument against the internalist approach to knowledge requires a couple of assumptions which I am not going to defend in this section. This is why my treatment of the vertical epistemic regress problem will be developed at the more formal level of such discussions as can be found in Bonjour (1980, 54 f.), Van Cleve (1984, 563), Dancy (1985, 129 ff.) and Alston (1989, 210 f.).

The internalist version of the JTB account of knowledge introduced in section 3.1 will provide a useful illustration:

S knows that *p* **iff** 1) *p* is true
2) S believes that *p*
3) S is justified in believing that *p*

It must be stressed that the argument does not apply to internalist accounts of knowing generally. For example, it does not apply to the internalist construal of reliabilism envisaged in section 3.1 if the transparency demand of internalism is met by requiring that S be merely *externally* justified in believing that the conditions that are to be fulfilled for her to know that *p* are indeed satisfied. However, the argument does apply to all mainstream internalist accounts of knowing in virtue of their requiring that the satisfaction of the conditions, *whatever they are*, that must be met for S to know that *p* be epistemically accessible to S.

Now, assuming that 'S has epistemic access to *p*' is construed as 'S is justified in believing that *p*', the JTB account of knowing is such that satisfaction of condition 3 entails epistemic access to the satisfaction of condition 1. Furthermore, there seems to be no doubt that S can have epistemic access to the satisfaction of condition 2. The vertical epistemic regress problem arises with condition 3: to have epistemic access to the satisfaction of this condition, S must be justified in believing that she is justified in believing that *p*. Our internalist construal of the JTB account of knowledge thus turns out to require the satisfaction of a new condition for S to know that *p*, namely:

4) S is justified in believing that she is justified in believing that *p*

But then, S will need to have epistemic access to the satisfaction of condition 4 as well, because on an internalist conception of epistemic justification S cannot be justified in believing that she is justified in believing that *p* without being justified in believing that she is justified in believing that she is justified in believing that *p*. And this means that S cannot know that *p* unless yet another condition is satisfied, namely:

5) S is justified in believing that she is justified in believing that she is justified in believing that *p*

An infinite regress then appears to be unavoidable (one can easily see that things get even worse if we replace 'S is justified in believing that *p*' with 'S knows that *p*' as our interpretation of 'S has epistemic access to *p*' in the formulation of the internalist requirement upon knowledge).

A way to stop the vertical epistemic regress without adopting a full-blown externalist position might be to adopt a weaker form of internalism, as Dancy (1985, 133 ff.) has suggested. Dancy's brand of internalism does not equate the required epistemic access to the satisfaction of the conditions for knowing with knowledge or justified belief, but merely with (actual) belief. Accordingly, what is required from S is not that she know or justifiedly believe that the relevant conditions for her to know that *p* are satisfied, but merely that she *believe* that they are. Condition (4) is thus replaced by condition (4'): S believes that she is justified in believing that *p*.

Dancy's proposal avoids the regress, but for all of his assurances I cannot see why it should count as an internalist conception of knowledge. Mere belief is evidently insufficient for epistemic access. Therefore it comes as no surprise that the only way in which Dancy's proposal captures the intuitions that appear to motivate epistemological internalism is that it rules out that sense of knowledge in which we ascribe knowledge to infants and all kinds of organisms which are well-adapted to their environment. Moreover, it rules out any kind of knowledge the cognitive subject is not aware of, such as certain kinds of behavioural dispositions and unconscious expectations. Apart from these features, whatever plausibility Dancy succeeds in providing for the claim that his proposal is in any relevant sense internalist seems to me to come from an equivocation.

This equivocation can be detected in Dancy's argument that epistemological internalism is fully captured by condition (4') *because* 'all that we can ask of a [person] is that [s]he retain beliefs which, so far as [s]he can tell, meet the conditions for justification' (Dancy 1985, 133). Now, what 'S believes that she is justified in believing that *p*' means is just that S *happens* to believe that she is justified in believing that *p*. That we could charge of (cognitive) irrationality a person that retained beliefs she believed to be unjustified is of course true, but is this everything there is to epistemological internalism? Does 'S believes that she is justified in believing that *p*' also mean that *p*, 'so far as [s]he can tell', meets the conditions for justification? That depends on how we construe the clause 'so far as [s]he can tell'.

Suppose Peter is pathologically jealous of his (faithful) wife; when he is sober, he (unjustifiedly) believes her to be unfaithful. But, oddly enough, when he is drunk he changes his mind and believes her to be faithful. Indeed, we can suppose that Peter's (intermittent) belief that his wife is faithful is justified, and that when he is under the effects of alcohol he also believes that it is. Under such circumstances, I very much doubt that saying that Peter *knows* that his wife is faithful when he is drunk would be consistent with the intuitions that motivate epistemological internalism. This is because condition (4') can be seen as capturing the idea that 'all that we can ask of a [person] is that [s]he retain beliefs which, so far as [s]he can tell, meet the conditions for justification' only if our construal of the clause 'so far as [s]he can tell' is so weak as not to impose any rationality requirement at all on S's belief-forming processes. But the idea that 'all that we can ask of a [person] is that [s]he retain beliefs which, so far as [s]he can tell, meet the conditions for justification' can be taken to express genuinely internalist intuitions only if the clause 'so far as [s]he can tell' is construed as implying that S ought to do her (cognitive) best in the assessment of the epistemic status of her own beliefs.

But it might be retorted that Peter's case represents a counterexample to the claim that Dancy's condition (4') successfully captures the intuititions that motivate epistemological internalism only on a rather weak account of epistemic justification. For S cannot be taken to be epistemically justified in believing that *p* merely because her evidence entails (or makes it probable) that *p* is true; if S is to be epistemically justified in believing that *p*, she must actually *infer* that *p* from the evidence in her possession. And since the reason why Peter, when drunk, thinks to be justified in believing that his wife is faithful is unlikely to be that he is better at drawing inferences when he is drunk than when he is sober, the hypothesis that his (intermittent) belief that his wife is faithful is epistemically justified seems utterly implausible.

This rejoinder will not do. Let us grant that S cannot be epistemically justified in believing that *p* unless she does infer that *p* from the evidence in her possession. Even so, S may *believe* that she is justified in believing that *p* for reasons that have nothing to do with the pursuit of knowledge.

Suppose there is an after-life. And suppose that Mrs. Smith's belief that there is an after-life is epistemically justified in the sense that it is both supported and inferred by the evidence in her possession. Then, if Mrs. Smith believes that her belief that there

is an after-life is epistemically justified, Dancy's condition (4') will be satisfied. However, we may also suppose that if Mrs. Smith's evidence for the existence of an after-life had been slightly defective, she would still have thought to be justified in believing that it existed — she would have done so, we may imagine, because she is consumed by the longing to meet again her untimely deceased husband. If so, it is hard to resist the impression that Mrs. Smith is in fact led to believe that she is justified in believing that there is an after-life by the 'wrong' kind of reasons, i.e., by reasons that have nothing to do with the pursuit of knowledge. But then, the mere fact that her belief that there is an after-life satisfies Dancy's condition (4') as well as the usual conditions of the JTB account of knowing cannot be taken to mean that she *knows* that there is an after-life in any genuinely internalist sense. Thus it seems reasonable to conclude that Dancy's condition (4') cannot be seen as doing justice to the internalist intuitions it is advertised as capturing. In so far as Dancy's requirement that S believe to be justified in believing that *p* does appear to do justice to those intuitions, it is merely because by introducing the clause 'so far as [s]he can tell' he is smuggling in an assumption about the rationality of S's belief-forming processes which is certainly not captured by (4').

The only way of stopping the vertical epistemic regress without giving up knowledge-internalism is, I think, to deny that an internalist conception of knowledge must involve an internalist conception of epistemic justification. In other words, one ought to maintain both (1) that S's knowing that *p* involves S's having epistemic access to the satisfaction of the conditions that must be met in order for her to know that *p*, and (2) that S can be justified in believing that *p* without having epistemic access to the satisfaction of the conditions that must be met in order for her to be justified in believing that *p*. S's having epistemic access to the satisfaction of condition (4) will then fail to be a necessary condition of her having epistemic access to the satisfaction of condition (3) — more generally, S's having epistemic access to the satisfaction of condition n + 1 will fail to be a necessary condition of her having epistemic access to the satisfaction of condition n, for any $n \geq 3$ — and the vertical epistemic regress will be avoided.

However, this way of avoiding the vertical epistemic regress is not without problems for the upholder of an internalist epistemology. First of all, it has the consequence that if S is to know that *p*, at some stage she will be prevented from knowing that she knows that she knows... that *p*. This contradicts both the objectivity and the infallibility principles. But while a non-dogmatic epistemologist can easily do

without the latter principle, rejecting the former seems hardly compatible with any kind of internalist account of knowing. Secondly, the acceptance of justification-externalism would commit the upholder of knowledge-internalism to provide a plausible account of why S's knowing that *p* should entail that, while epistemic justification must be transparent to S at the ground-level, it must cease to be so at higher levels. For in the absence of such an account, the claim that epistemic justification must at some level (higher than the ground level) cease to be transparent would seem definitely *ad hoc*.

I am not saying that these problems are insuperable — the domain of the objectivity principle might be suitably restricted and different sorts of epistemic justification might be cleverly distinguished. However, if denying that epistemic justification must in general be transparent to the believing subject is the only way in which the upholder of an internalist view of knowledge can avoid the vertical epistemic regress, at the very least she will have to reconcile herself to the fact that *acceptance of knowledge-internalism involves rejection of justification-internalism*. The full significance of this point will become clear in chapter 5, where it will be argued that after the decline of dogmatic epistemologies the strongest motivation for accepting an internalist view of knowledge is given in fact by the need of an internalist view of *epistemic justification*. The reason that prompts a number of non-dogmatic epistemologists to subscribe to knowledge-internalism is that they still believe (perhaps as an effect of their Cartesian legacy) that endorsing an internalist view of knowledge is the only way of endorsing an internalist view of epistemic justification. But if this is so, then even if the vertical epistemic regress problem cannot be harnessed to devise a strict refutation of knowledge-internalism, it can be harnessed to show that an internalist view of knowledge involves commitments which are inconsistent with the most likely motivation of the adoption of knowledge-internalism. And this should lead many supporters of justification-internalism to give up a position which is in fact incompatible with their deeper convictions.

An Externalist View of Knowledge

We saw in the last chapter how the internalist conception of knowledge outlived the dogmatic epistemologies by which it had been originally motivated. However, the search for a Gettier-proof 'third condition' for knowledge, the emergence of such problems as have been discussed in section 3.4, and a less intellectualized view of our cognitive processes have recently led to the investigation of less restrictive models of knowledge. People like David Armstrong, Fred Dretske, Alvin Goldman, and Robert Nozick have been trying to develop, in the last two decades or so, a thoroughly externalist alternative to the traditional accounts of knowing.

This chapter contains an investigation of two of their proposals. The focus is on accounts which make room (or which can be so modified as to make room) for the possibility that human beings have, at least in certain circumstances, knowledge of their knowledge. We saw in the last chapter that the rejection of the internalist project of modern epistemology need not be regarded as the denial of any possibility of access to one's own epistemic situation. What an externalist analysis of knowing is committed to deny is rather the necessity that each and every instance of knowledge allow such an access in order to qualify as knowledge. The point of the externalist analyses of knowing I am interested in is, ultimately, that there can be genuine instances of knowledge which are not arrived at through the sort of methods favoured by post-Cartesian epistemologists, and that such instances cannot be simply dismissed as invalid if one wants to stand any chance of improving one's own epistemic situation. The aim of this chapter is thus to outline the main features of the sort of externalist account of knowing

on the basis of which correspondence truth can be successfully argued to be epistemically at least as accessible as epistemic truth.

Sections 4.1 and 4.2 discuss Robert Nozick's 'tracking' analysis of (factual) knowledge, while sections 4.3 to 4.5 discuss Fred Dretske's 'information-theoretic' analysis of (perceptual) knowledge. Both analyses display 'naturalistic' features, focus on the discriminatory aspects of cognitive processes, and support the so-called 'relevant alternatives' answer to scepticism. Nozick's analysis provides definite truth-conditions for 'S knows that she knows that *p*' and has the nice consequence that those conditions are often fulfilled, without supporting any sort of objectivity or infallibility principle. Unfortunately, it requires a questionable possible-worlds semantics for counterfactual conditionals and it is haunted by compelling counterexamples showing that it fails to provide an adequate analysis of our pre-philosophical intuitions about knowledge. Dretske's analysis, on the other hand, is more limited in scope, suffers (at least in its original form) from a certain ambiguity, and does not provide truth-conditions for 'S knows that she knows that *p*'. But it seems to have no counterintuitive consequences, and one can think of completing it with further elements to cover a wider range of cognitive phenomena. I shall discuss how it can be extended to cover knowledge of natural laws and knowledge of one's knowledge. The admittedly sketchy picture of knowledge which will ensue from this discussion will provide the background for addressing, in chapter 5, the role of epistemic justification in our argumentative practices.

4.1 Nozick's Tracking Analysis of Knowing

Nozick's tracking (or counterfactual) analysis of 'S knows that *p*' (see Nozick 1981, ch. 3, esp. 172-178; 197-211) is supposedly immune to the Gettier (1963) counterexamples to the JTB analysis of knowing, and promises to provide a fresh treatment of the sceptic's denial of the possibility of human knowledge of the external world.

Nozick's analysis of knowledge is as follows[1]:

[1] I give the substance of Nozick's analysis. In his (1981, 179-196), Nozick introduces some further refinements which we need not take into consideration here.

S knows that p **iff**
1) p is true
2) S believes that p
3) $\neg p \; \Box\!\!\rightarrow \neg$(S believes that p)
4) $p \; \Box\!\!\rightarrow$ S believes that p

Conditions 1 and 2 need no comment. Conditions 3 and 4 are intended to capture the non-accidentality of S's true belief that p without appealing to the traditional concept of epistemic justification. The box-arrow stands for the subjunctive conditional 'If it were the case that..., then it would be the case that...'. Nozick seems to suggest that the subjunctive conditionals occurring in his definition should be taken at face value. In fact, he would like to claim that his definition does not commit him to any particular semantics for subjunctive conditionals. But he has to refer to Robert Stalnaker's and David Lewis' 'possible-worlds' analyses as to the best candidates for such a role (see Nozick 1981, 680, note 8), because there are relevant circumstances in which no informal reading of condition 4 seems to be possible.

If conditions 1 and 2 are met, condition 3 can be easily read without any 'possible-world' aid as 'If it were the case that not-p, S would not believe that p'. It is more difficult to provide an intuitively satisfactory reading of condition 4. For if it is actually the case that p and S does believe that p, then the subjunctive conditional 'If it were the case that p, S would believe that p' does not appear to make any sense (because in ordinary English the use of a subjunctive conditional requires that the hypothesis described in its antecedent be not assumed to be realized). According to a suitable revision of Stalnaker's and Lewis' possible-worlds semantics, if one wants to decide whether the subjunctive conditional $p \; \Box\!\!\rightarrow \; q$ is true, one will have to examine those possible p-worlds that are closest (most similar) to the actual world, and see if q holds true in all of them. If it does, the subjunctive conditional holds true in the actual world. Thus condition 4 turns out to mean something like: 'In all those worlds in which p holds true that are closest to the actual world, it is also true that S believes that p'. If, under the assumption that conditions 1 and 2 are met, one wanted to render condition 4 in a slightly more familiar language, she would probably have to go for something like this: '(p is true and S believes that p, and) S would believe that p in all those (slightly) different circumstances in which p were to hold true'.

The idea that Nozick's counterfactual analysis of knowing is meant to capture is that S knows that p only if her belief that p is 'sensitive' both to the truth and to the

falsity of *p*. If S knows that *p*, she does not merely happen to believe truly that *p*: 'To know that *p* is to be someone who would believe it if it were true, and who wouldn't believe it if it were false' (Nozick 1981, 178). Nozick sums up this idea of a subjunctive connection between S's belief about *p* and *p*'s truth-value by saying that S's belief 'tracks' the truth that *p*:

> To know is to have a belief that tracks the truth. Knowledge is a particular way of being connected to the world, having a specific real factual connection to the world: tracking it.
>
> (Nozick 1981, 178)

The tracking analysis belongs to the reliabilist family, in so far as the appeal to the counterfactual dependency of S's belief that *p* on the relevant states of the world is meant to spell out the idea that the truth of S's belief that *p* is not accidental, but ensues from the use of a reliable cognitive method, hence representing a faithful indicator of the actual state of the world. Nozick himself emphasizes the similarity between his analysis and such reliabilist accounts of knowing as those proposed by David Armstrong, Fred Dretske and Alvin Goldman (see Nozick 1981, 689 f.).

We can now turn to the relevance of Nozick's analysis for the sceptic's denial of the possibility of human knowledge of the 'external' world.

Nozick addresses a sceptical challenge of the kind described in sections 2.2 and 2.3: if one were a brain in a vat on Alpha Centauri being properly stimulated by some super-psychologists, one would not know what one usually assumes to know. But then, how can one possibly have any knowledge of the world at all, if one cannot rule out the possibility that one is just a brain in a vat on Alpha Centauri being systematically deceived about one's own situation[2]?

This sceptical challenge relies on the principle P that knowledge is closed under known logical implications (Nozick 1981, 204 ff.). This principle says, roughly, that if S knows that *p* and if S knows that '*p* entails *q*', then S also knows that *q*. Principle P enables the sceptic to argue, by *modus tollens*, that if S knows that '*p* entails *q*' and she does not know that *q*, then she cannot possibly know that *p*.

---

[2] Unlike Putnam, Nozick does not assume that absolutely everybody is a brain in a vat. Nozick's anti-sceptical argument is not meant to be a 'transcendental' argument.

Nozick notes that it is possible to argue over the details of principle P. For example, S can know that *p* and also that '*p* entails *q*'; but if for some reason she fails to draw the inference to *q* (e.g., because she does not know that she knows that *p*, or because she does not know that she knows that '*p* entails *q*'), she may well fail to know that *q*. However, Nozick points out that it won't be by quibbling over the details of P that one will successfully rebut the sceptical challenge. Principle P does not need minor amendments: it must be flatly rejected.

Let *p* be 'S is sitting in her room reading a book' and *q* be 'S is not a brain in a vat on Alpha Centauri'. We can suppose that S knows that '*p* entails *q*' (i.e., that she cannot both be sitting in her room reading a book and be a brain in a vat on Alpha Centauri). Yet S does not seem to know that *q* (all evidence in her possession is compatible with her being a brain in a vat on Alpha Centauri). So how can S possibly know that *p* (i.e., that she is sitting in a room reading a book) if she does not know that *q* (i.e., that she is not a brain in a vat on Alpha Centauri)? The sceptical predicament arises from the possibility of imagining a whole set of 'sceptical worlds' supposedly different from the one in which we believe we live, but which nevertheless would cause us to have qualitatively identical experiences to the ones we actually happen to have. Such 'sceptical worlds' would be *evidentially indistinguishable* from, and hence *doxically identical* to, the one which is supposedly ours (different worlds are said by Nozick to be 'doxically identical' for S iff S would have exactly the same beliefs in any of them). But, so goes the sceptical argument, if we cannot tell that we are not living in one of those sceptical worlds, we cannot know anything at all about our own world. For none of our actual beliefs (about the world) would be true if we were in fact living in a sceptical world.

However, Nozick's analysis of 'S knows that *p*' has the nice consequence that in general S can be said to know that *p* and that '*p* entails *q*' without being required to know that *q*. In other words, Nozick's analysis entails that knowledge is *not* closed under known logical implications (Nozick 1981, 204-211).

Let us see how the tracking analysis of knowing handles our example involving the propositions *p*, 'S is sitting in her room reading a book', and *q*, 'S is not a brain in a vat on Alpha Centauri'. It is immediately clear that S cannot be said to know that *q*, because if she were a brain in a vat on Alpha Centauri ($\neg q$) she would nevertheless believe that she were not (that is, she would believe that *q*), since that sceptical world

would be doxically identical to what she takes to be the actual world. So Nozick will grant the sceptic that S does not know that she is not a brain in a vat on Alpha Centauri. Yet, if S tracks the truth that *p*, she can be said to *know* that she is sitting in her room reading a book:

> For (3') [i.e., 'if *q* were false, S wouldn't believe that *q*'] talks of what S would believe if *q* were false, and this may be a very different situation from the one that would hold if *p* were false, even though *p* entails *q*. [...] There is no reason to assume the (closest) not-*p* world and the (closest) not-*q* world are doxically identical for you, and no reason to assume, even though *p* entails *q*, that your beliefs in one of these worlds would be a (proper) subset of your beliefs in the other.
>
> (Nozick 1981, 206 f.)

In other terms, when we assess the truth-value of 'if *p* were false, S wouldn't believe that *p*', we need to consider those possible not-*p* worlds that are closest (most similar) to the actual world, and see if 'S does not believe that *p*' holds true in all of them. We do not have to consider those possible not-*p* worlds which, like all not-*q* worlds, are most distant from (most dissimilar to) the actual world. Thus, since none of the relevant not-*p* worlds will be a world in which S falsely believes that *p*, condition 3 will be satisfied, and S can be said to track the truth that *p*, even if she cannot be said to track the truth that *q*.

This result should not be overestimated. As Craig (1989) has pointed out, Nozick's analysis entails that, if any sceptical world is a close possible world, we do not really know what we think we know. But since the actual world *is* a close world, if we want to defeat the sceptic we need to be in a position to assert that the actual world is *not* a sceptical world. However, if we are in such a position, then the sceptic must already have been defeated without recourse to the tracking analysis; and if we are not, the tracking analysis will not help. Nozick may thus be right in his contention that principle P is false, but that does not show that we can establish that we have any genuine knowledge of the world.

Craig's argument is valid, but it does not undermine Nozick's actual claims. Craig's argument shows that the tracking analysis of knowing is unable to *prove* that we have any actual knowledge of the world. But that has nothing to do with what Nozick's analysis is intended to prove in the first place. Nozick's analysis is intended to prove that

S may (not: does) know that *p* even without knowing that she is not a brain in a vat on Alpha Centauri. In other terms, the tracking analysis of knowing entails that *knowing that one's world is not a sceptical world is not a necessary condition for having any knowledge at all.* Craig's reply does not undermine this result, which is a way to express what is now known as the 'relevant alternatives' answer to the sceptic. This argues that the sceptic's requirement that one be able to prove beyond any possible doubt the existence of actual instances of knowledge is unjustified because in order to know that *p*, one must merely be able to tell the fact that *p* from those possible states of affairs that constitute *relevant* alternatives to *p*, where the set of relevant alternatives one must be able to exclude is merely a subset of *all* possible alternatives to *p* (Dretske 1981b, esp. 371 f.).

Why should this result have any bearing on the sceptical challenge though? Are we epistemologically any better off after Nozick has told us that S *might* be in a position to know that she is sitting in her room reading a book, although she cannot prove that she is? I think this question can be answered in the affirmative, because it is not correct to say that Nozick's analysis merely tells us that S *might* be in a position to know that she is sitting in her room reading a book. The reason why this is so is that Nozick's analysis meets the requirement that an adequate externalist account of knowing should allow people to know that they know this or that. Nozick's analysis tells us that, if S is indeed sitting in her room reading a book and knowing that she is, she can normally *know* that she knows that she is. This is immediately clear as soon as one pays attention to the fact that, if S's knowing that *p* is S's truly believing that *p* and having her belief track the truth that *p*, S's knowing that she knows that *p* will be S's truly believing that she knows that *p* and having her belief track the truth that she knows that *p*. And there is no more reason to believe that the conceivability of sceptical worlds should undermine the possibility of S's knowledge of her knowledge that *p*, than there is to believe that it should undermine the possibility of S's knowledge that *p*:

> The skeptic's doxically identical possibility SK [that S might be a brain in a vat on Alpha Centauri] did not show the falsity of condition 3, not-*p* □→ not-(S believes that *p*). For even if *p* were false his possibility SK wouldn't hold, anyway. Similarly his possibility will not show the falsity of
>
> not-3 □→ not-(S believes that 3)

when even if 3 were false his possibility SK also wouldn't hold; that is, when not-3 $\Box\!\!\rightarrow$ SK does not hold.

(Nozick 1981, 246 f.)

As before, when we assess the truth-value of 'if S did not know that *p*, S wouldn't believe that she knew that *p*', we have to consider those possible not-(S knows that *p*) worlds that are closest (most similar) to the actual world, and see if 'S doesn't believe that she knows that *p*' holds true in all of them. We do not have to consider those possible not-(S knows that *p*) worlds which, like all SK worlds, are most distant from (most dissimilar to) the actual world.

If nevertheless one still gets the impression that S cannot *know* that she knows that she is in her room reading a book unless she can rule out the possibility that her world is a sceptical world, this must be because one is still under the spell of the quest for certainty, with its suggestion that one cannot know that *p* (e.g., that one knows that *q*) unless one can prove beyond any possible doubt that *p*. But if one accepts the tracking analysis of knowledge, it will be a factual matter whether S, besides knowing that *p*, also knows that she knows that *p*:

> If knowledge is a real relationship in the world, such as tracking, then it will be a fact that you stand in that relationship to *p*; so room will be left for failing to stand in that very (tracking) relationship to the fact that you stand in it to *p*.
>
> (Nozick 1981, 246)

However, in ordinary circumstances there is no reason to believe that the mere possibility of sceptical worlds should prevent S from knowing that she knows that she is in her room reading a book[3].

The price to be paid for these results is of course the surrender of those features which are usually associated with internalist, and especially dogmatic, accounts of knowing. When S knows that *p*, she may well know, as a matter of fact, that she knows that *p*, but the requirement that she do will no longer be embodied in the concept of

---

[3] What the conceivability of sceptical worlds *may* show, is that there must be some limit to the iteration of the knowledge operator K: perhaps S cannot know that she knows... that she knows that p. S will not know that she knows... that she knows that p when one of the not-(S knows that S knows... that S knows that p) worlds that are closest (most similar) to the actual world is indeed a sceptical world (see Nozick 1981, 694 f.).

knowledge. Accordingly, the tracking account of knowledge will not support any objectivity, let alone infallibility, principle. Nor will it support the requirement that only epistemically justified beliefs be treated as legitimate candidates to the status of knowledge.

Summing up: the results of Nozick's analysis of knowledge are indeed relevant to the sceptical challenge, but not in the sense that they will enable someone claiming that *p* to prove beyond any conceivable doubt that she knows that *p*. Rather, the results of Nozick's analysis of knowledge are relevant to the sceptical challenge because they will enable someone claiming that *p* to argue against the sceptic that being able to provide such a proof is not a necessary condition for knowing that *p*. Therefore, one must not construe Nozick's argument as if it were designed to meet the sceptical challenge by proving the existence of instances of knowledge which satisfy standard internalist constraints. Nozick denies the very adequacy of those internalist constraints by showing that there is a sense in which we may *know* the world (i.e., have a 'specific real factual connection' to it), and indeed know that we do, even if those constraints are not satisfied. Accordingly, it will not help to reject his argument merely by showing that it does not prove the existence of instances of knowledge satisfying the internalist requirements on knowing. What will have to be discussed is rather whether an externalist view of knowledge can provide a satisfactory alternative to the internalist orthodoxy, or whether externalism must be seen, on the contrary, 'as simply abandoning the traditional idea of epistemic justification or rationality and along with it anything resembling the traditional conception of knowledge' (Bonjour 1980, 70). In the next section I shall then start to investigate how Nozick's tracking account of knowing fares as an *analysis* of our ordinary intuitions about knowledge.

4.2 The Flaws of the Tracking Analysis of Knowing

Post-Gettier epistemological discussions are full of examples designed to capture our intuitions about knowledge in order to refute this or that analysis of knowing. In fact, many of these examples do not succeed in capturing intuitions shared by all the participants in the game, not to mention ordinary, non-philosophical folks. This is surely true of some alleged counterexamples to the tracking analysis of knowing. The first three

counterexamples I shall introduce in this section I take to be uncontroversial. But the last two, which may be seen as representative of a larger class of purported counterexamples to the externalist analyses of knowing generally, do not seem to me to express genuine, untutored intuitions about knowledge.

The first counterexample I want to present is due to Risto Hilpinen:

> I am looking at a thermometer that is accurate within the range of 0° to 100°. At all temperatures below 0° the thermometer registers 0°. By observing its reading of 70° I come to believe that it is not —50°.
>
> (Alan Goldman 1988, 58)

It seems clear that I *know* that it is not —50°. But if it were, I would not believe it to be. So condition 3 of the tracking analysis is not satisfied, although I clearly seem to know that it is not —50°. As Alan Goldman points out, the use of a measuring instrument with a limited range can produce knowledge of facts whose failure cannot be tracked by the knowing subject, in virtue of the distance separating the closest not-*p* world from the actual world. This fact provides a method for devising a whole class of counterexamples showing that Nozick's analysis is too strong. And I do not know of any easy way in which the tracking account could be revised to handle such situations.

But Nozick's analysis can also be shown to be too weak. The following counterexample is due to Colin McGinn:

> You visit a hitherto unexplored country in which the inhabitants have the custom of simulating being in pain. You do not know that their pain behavior is mere pretence, and so you form the belief of each person you meet that he or she is in pain; imagine you have acquired a great many false beliefs in this way. There is, however, one person in this country who is an exception to the custom of pain pretence: this hapless individual *is* in constant pain and shows it [...]. You also believe of this person, call him *N*, that he is in pain.
>
> (McGinn 1984, 532 f.)

It seems clear that the belief that *N* is in pain cannot count as a true instance of knowledge, because it is not the outcome of an ability to discriminate between people who are really in pain and people who are not: its truth is in some relevant sense accidental. Nevertheless, the belief that *N* is in pain does satisfy Nozick's conditions for knowledge (in particular, if *N* were not in pain, you would not believe that he was).

The pain-simulators counterexample arguably involves knowledge inferred from evidence. But the same point can be made with reference to perceptual knowledge:

> Suppose you are surrounded by straight sticks immersed in water that therefore look bent to you; you, however, take them to be in air, and so you falsely believe of each of them that it is (really) bent. There is, though, one stick that is *not* immersed in water and it really *is* bent; on the basis of how that exceptional stick looks you believe it to be bent.
>
> (McGinn 1984, 533)

Nozick's analysis implies that you know that the exceptional stick is bent. But again, your belief that it is appears to be only accidentally true, for you also believe that the straight sticks are bent. And this again goes to prove that Nozick's analysis is too weak.

These are the counterexamples I take to be uncontroversial, and I think they suffice to show that Nozick's analysis is seriously flawed. The following counterexamples, on the other hand, do not seem to me to be grounded on uncontroversial intuitions.

The first one is a modified version of a Shope's example and is discussed by Nozick himself (see Nozick 1981, 190). Imagine a machine which displays, in alternate time periods, a hologram of a vase. Being activated by a vase-detector, the machine displays the hologram only when a vase is in its box. By seeing the hologram, a person comes to believe that there is a vase before her (which is true). Since the machine displays the hologram just in case it detects the presence of a vase, conditions 3 and 4 of the tracking analysis are satisfied, and Nozick is forced to say that the person *knows* that there is a vase before her. This consequence seems to be counterintuitive, but Nozick sticks to his analysis by noting that it would be wrong to hold that a person knows that *p* only if she has no false beliefs about the process via which she came to believe that *p*. After all, the 'Greeks had many false beliefs about the visual process' (Nozick 1981, 190), but nobody would think of denying that they were capable of having visual knowledge of their surroundings!

Another counterexample which makes a similar point against the externalist analyses of knowing is due to Laurence Bonjour:

> Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a

> cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power, under circumstances in which it is completely reliable.
>
> (Bonjour 1980, 62)

All the relevant conditionals come out true, and anyone endorsing Nozick's analysis will be committed to claim that Norman knows that the President is in New York City. This again may seem to be counterintuitive, because Norman is not in a position to defend the rationality of his believing that the President is in New York City.

Both the hologram and the clairvoyance counterexamples make the point that epistemic justification is an essential ingredient of knowledge. They can be taken as representative of a larger set of fundamentally interchangeable counterexamples[4], the purpose of which is not just to undermine Nozick's tracking analysis, but to reject the naturalistic approach to knowledge exemplified by this analysis. I say 'naturalistic' rather than 'externalist' because there are externalist analyses of knowledge which are not even apparently affected by such counterexamples. This is because the conclusion these conterexamples are designed to prompt is merely that a true belief exhibiting the 'right' factual connection to the world but the possessor of which is uncapable of justifying cannot be counted as a genuine instance of knowledge. But an externalist analysis of 'S knows that *p*' which required not just that S's true belief that *p* exhibit the 'right' factual connection to the world, but also that S be *justified* in believing that *p*, would be obviously immune to these counterexamples (more about this in the next chapter). It must be emphasized that the sort of epistemic justification that is alleged to be essential for knowledge is of course *internal* epistemic justification. The cognitive subjects in both the hologram and the clairvoyance counterexamples lack the capacity of providing an (adequate) *internal* justification for their (true) beliefs. This is why the point made by these counterexamples does not merely affect Nozick's tracking analysis, but any account

[4] See, for example, Lehrer's (1990, 163 f.) Mr. Truetemp example, whose protagonist is unaware of having a 'tempucomp' ('a small device which is both a very accurate thermometer and a computational device capable of generating thoughts') implanted in his head. Mr. Truetemp's beliefs about the temperature track the truth, but he accepts them unreflectively and has no idea whether they are reliable or not. Shall we say that he *knows* that the temperature is 104 degrees Fahrenheit when he thinks it is?

of knowing that makes of epistemic justification either an unessential or a potentially inaccessible (to the knowing subject) feature of knowledge[5].

I think that these counterexamples do not express genuine intuitions about *knowledge*, but merely an internalist (and I believe misguided) reconstruction of the fact that S's knowing that *p* does not licence, as such, S's *claiming* to know that she does. However, I shall postpone until chapter 5 a detailed discussion of this matter, because in sections 4.3 to 4.5 I wish to present and discuss an externalist analysis of knowing which provides a correct treatment of (what I take to be) the actual counterexamples to Nozick's analysis while retaining its most attractive features.

4.3 Dretske's Information-Theoretic Analysis of Knowing

Commenting on the failure of Nozick's analysis to provide *sufficient* conditions for knowledge, McGinn (1984, 536 f.) points out that the tracking analysis delivers the wrong results because it spells out the reliabilist insight in terms of a unique proposition *p* (along with S's disposition to believe it), without taking into account the reliability of S's dispositions to believe a larger set of propositions which are also relevant to whether S knows that *p* or not. The failure to take into account these further propositions accounts for the incapacity of Nozick's analysis to capture the idea that knowledge involves the ability to *discriminate* truth from falsehood within a range of relevant alternatives. In the bent stick case, for example, S is not capable of discriminating the truth that a particular stick is bent from the relevant alternative that it might be a straight stick immersed in water (that this should be counted as a *relevant* alternative is shown, in the case at hand, by the fact that S is supposedly surrounded by straight sticks looking bent precisely because they are immersed in water).

McGinn suggests that an externalist analysis of knowledge may be successfully provided by a *discrimination* account, as opposed to a tracking one.

Discrimination accounts of knowing focus on the discriminatory powers of the cognitive subject. According to such accounts, it is not merely the counterfactual dependency of S's belief that *p* on the fact that *p* that turns S's belief into knowledge.

---

[5] The clairvoyance counterexample was in fact originally intended as a refutation of Armstrong's (1973) reliabilist analysis of knowledge.

A satisfactory analysis of S's knowing that *p* must refer to S's capacity to *distinguish* or *discriminate* the actual state of affairs in which *p* is true from all *relevant* possible state of affairs in which *p* is false, where the individuation of these relevant alternatives depends, at least in part, on which *other* propositions are true or false in S's epistemic situation. And in spite of what may seem at first glance, Nozick's analysis lacks the resources for a correct individuation of the relevant alternatives to *p*'s truth.

An externalist account of knowing which could be described, in McGinn's terms, as a *discrimination* account is Fred Dretske's information-theoretic analysis of '*K* knows that *s* is *F*'. Dretske (1981a, 86) defines '*K* knows that *s* is *F*' as '*K*'s belief that *s* is *F* is caused (or causally sustained) by the information that *s* is *F*'.

The concept of 'information' that appears in Dretske's definition must be construed in a slightly revised information-theoretic sense. This is how Dretske defines 'informational content':

> *Informational content*: A signal *r* carries the information that *s* is *F* = The conditional probability of *s*'s being *F*, given *r* (and *k*), is 1 (but, given *k* alone, less than 1)
>
> (Dretske 1981a, 65)

The parenthetical *k* stands for the knowledge the receiver already possesses (if anything) about the source. For example, if one already knows that *s* is either red or blue, a signal that eliminates the possibility of *s*'s being blue carries the information that *s* is red, in so far as it raises the conditional probability of *s*'s being red, given *k*, from 0.5 to 1 (Dretske 1981a, 65). The fact that *k* appears in the definition of the informational content of a signal does not make Dretske's definition of knowledge circular (*k* does not contain the information that *s* is *F*), but recursive.

Saying that '*K*'s belief that *s* is *F* is caused (or causally sustained) by the information that *s* is *F*' means that it is caused (or causally sustained) by a physical signal carrying the information that *s* is *F*. Dretske emphasizes that the information that *s* is *F* can be said to cause *K*'s belief that *s* is *F* only if the physical signal carrying the information that *s* is *F* carries that information in virtue of having the property *F'* and if it is precisely in virtue of the signal's being *F'* that it causes *K*'s belief. For example, the spy knows that the courier has arrived because she hears three quick knocks on the door followed by a pause and then by another three quick knocks. The feature that

carries the information that the courier has arrived is not the amplitude or pitch of the sounds, but their sequence. The spy's belief that the courier has arrived is caused by her hearing that particular sequence, not by any other feature of the knocks (Dretske 1981a, 87).

Dretske's definition allows that *K*'s belief that *s* is *F* be 'causally sustained', rather than caused, by the information that *s* is *F*, because a true belief produced by unreliable means can nevertheless qualify as a genuine instance of knowledge if it is subsequently 'causally sustained' by the relevant information (Dretske 1981a, 88-90).

Dretske's information-theoretic account of knowing is exposed to the very same internalist objection (as expressed by the hologram and clairvoyance counterexamples) raised against Nozick's tracking account, but it does avoid the other, more pertinent, counterexamples.

According to Dretske's definition of knowledge, the thermometer reading of 70° does carry the information that it is not —50° (Dretske would say that the information that it is not —50° is 'nested'[6] in the thermometer reading), thus turning my belief that it is not —50° into a genuine instance of knowledge. The fact that, if it were —50°, I would not believe it to be, does not prevent my present belief that it is not —50° from being caused by the information that it is not —50°.

That Dretske's information-theoretic analysis provides the intuitively right results in the cases of the pain-simulators and the bent stick is less straightforward but equally true (see Dretske 1981a, 123-134). This is because the very existence of a communication system presupposes a sharp distinction between a source *about which* information is received and a channel *over which* this information is received (Dretske 1981a, 133).

Dretske discusses the case of a voltmeter attached to a resistor in an electric circuit. If the voltage drop across the resistor is 7 volts, the pointer of the voltmeter will move up the instrument scale until it will come to rest on 7. If, for the sake of the example, it is assumed that the voltage across the resistor has only ten (equally) possible values, the information generated at the source will be approximately 3.3 bits[7].

---

[6] 'The information that *t* is *G* is nested in *s*'s being *F* = *s*'s being *F* carries the information that *t* is *G*' (Dretske 1981a, 71).

[7] The amount of information *I(s)* generated by a process *s* the result of which is the reduction of *n* equally likely possibilities to 1 is given by the formula: $I(s) = \log n$ (where log is the logarithm to the base 2).

> If [...] the instrument is functioning properly, the equivocation between source (voltage across the resistor) and receiver (position of pointer) is zero: the pointer's registering 7 volts carries the information *that* the voltage drop across the resistor is 7 volts. And this, according to the present account, is what enables the user to *tell* (know) what the voltage across the resistor is.
>
> (Dretske 1981a, 112 f.)

But what happens if the voltmeter is *not* functioning properly? The position of the pointer depends on many factors other than those associated with the source (e.g., the resistance of the leads connecting the instrument to the circuit, the calibration of the restraining spring attached to the armature to which the pointer is affixed, etc.). If any of these factors is changed, the 7-volt reading may be obtained even if the voltage drop across the resistor is not 7 volts. So how can the pointer reading carry the information that the voltage across the resistor is 7 volt if the same reading can be produced, for example, by a 5-volt drop *and* a weakened restraining spring? The conclusion of this tendentious line of reasoning is of course, as Dretske points out, that 'voltmeters and other measuring devices never deliver the information they are designed to deliver' (Dretske 1981a, 113), which is the information-theoretic analogue of the notorious sceptical thesis that our sensory experience, in so far as it depends, not only on the properties of the objects we perceive, but also on the condition of our sensory apparatus, on the nature of the illumination, etc., *is always equivocal.*

To this Dretske's rejoinder is that the channel of communication need not be *known* to be reliable in order to carry information about the source: it only needs to *be* reliable. *If* our voltmeter is in proper working order, it will deliver the information that the voltage across the resistor is 7 volts, whether we believe it to be reliable or not. Of course springs lose their elasticity over prolonged periods and wires break: therefore measuring instruments need periodic calibration and adjustment. But once the instrument has been initially calibrated and adjusted,

> repeated checking of the leads (to see whether they have changed their resistance), daily tests of the spring's elasticity (to see whether it has unpredictably changed its coefficient of elasticity), recounting the number of windings on the internal electromagnets (to make sure the same current flow will generate the same magnetic field) is unnecessary.
>
> (Dretske 1981a, 116)

Repeated checking may change our disposition to *trust* the readings of the instrument, it will not change the instrument's reliability. The point is, in order for communication to be at all possible, some conditions must be taken as fixed, they must be taken as providing no new information, they must qualify, in other words, 'as the framework *within which* communication takes place, not as a source *about which* communication takes place' (Dretske 1981a, 116). If the communication channel is indeed fixed and does not produce equivocation about the source, then the flow of information (and knowledge) will be possible; but the communication channel need not be *known* to be reliable in order to ensure the flow of information.

Whether a channel is reliable depends on which alternative states are irrelevant for the sake of the communication of information. We now get to the reason why Dretske's account gives the correct results in the pain-simulators and the bent stick cases. Let us take the pain-simulators example. Observation of pain behaviour is usually a reliable method to decide whether someone is in pain or not. If our hapless man lived in England, there would be indeed nothing counterintuitive in claiming that we could know, from the observation of his behaviour, that he is in pain. The fact that we cannot know that he is in pain in his own exotic country is due to his fellow-citizens' custom of simulating being in pain: this custom introduces a relevant alternative (which would be absent if *N* lived in England) which our cognitive method must be able to exclude if we are to have any knowledge of the fact that *N* is in pain. While observation of pain behaviour is a reliable method to detect pain in England, it is not in *N*'s country. Therefore, while the observation of pain behaviour in England does carry the information that somebody is in pain, it does not in *N*'s country. If the communication channel is 'that set of existing conditions that have no relevant alternative states, that in fact generate no (new) information' (Dretske 1981a, 123), what qualifies as a communication 'channel' in England does not qualify as a communication 'channel' in *N*'s country. The pain-simulation habit of *N*'s fellow-citizens makes it a relevant alternative that he might be merely simulating pain, and our observation of his pain behaviour can no longer carry the information that he *is* in pain, because it is no longer capable of excluding all the relevant alternatives to this state of affairs. (The bent stick case can be addressed along similar lines).

Having said this, it comes as no surprise that Dretske's information-theoretic analysis of knowing exhibits anti-sceptical consequences similar to those of Nozick's

analysis. *K*'s true belief that *s* is *F* can be caused (or causally sustained) by the information that *s* is *F* even if *K* does not know that she is not a brain in a vat on Alpha Centauri. A 'sceptical world' is not a relevant alternative to the actual world. Even though no conceivable signal can exclude the possibility that our world might be a sceptical world, this does not mean that all signals are genuinely equivocal (springs do lose their elasticity, but this does not mean that, if the voltmeter is in proper working order, we cannot know that the voltage across the resistor is 7 volts!). In other words, Dretske, like Nozick, denies the principle P (that knowledge is closed under known logical implications) and espouses the 'relevant alternatives' answer to the sceptic. On Dretske's analysis, the information that I am not a brain in a vat on Alpha Centauri is not 'nested' in the fact that I am sitting in my room reading a book: even if the former is a logical consequence of the latter, when I receive the piece of information that I am sitting in my room reading a book, the signal carrying this information does not exclude the (irrelevant) alternative that I might be a brain in a vat on Alpha Centauri, which means that I can know the former without knowing the latter.

Moreover, the particular version of the 'relevant alternatives' answer to the sceptic entailed by Dretske's analysis is not undermined by Kripke's unpublished example harnessed by Lehrer (1990, 183 f.) to argue that some version of principle P must be true, and hence that the 'relevant alternatives' answer to the sceptic must fail. As a matter of fact, Lehrer's argument only gets through on the supposition that the rejection of principle P is justified by appeal to a tracking analysis of knowing: it does not get through if it is justified by appeal to Dretske's information-theoretic analysis. Kripke's example is as follows: in a part of the country where a clever stage builder put up, unknown to me, red barn facades which look exactly like red barns, I see a blue barn, which is in fact the only real barn in the area, and I believe that I see a (blue) barn. Kripke points out that Nozick's tracking analysis has the unpalatable consequence that in this situation I would know that I see a blue barn without knowing that I see a barn (because I would not believe that I saw a blue barn if I did not see a blue barn, while I would believe that I saw a barn even if what I saw was merely a barn facade). So far, so good. But Lehrer's conclusion that 'if we try to escape from skepticism by rejecting the closure principle, we shall find ourselves committed to saying that we know that we see a blue barn when we do not know that we see a barn' does not follow. For Dretske's analysis of knowing does give, unlike Nozick's, the desired result that we

cannot know that we see a blue barn without knowing that we see a barn (because in the situation described by Kripke we cannot get the information that we see a blue barn without getting the information that we see a barn). Therefore it is not the rejection of principle P, but the tracking analysis of knowledge that involves commitment to the implausible claim that we might know that we see a blue barn without knowing that we see a barn.

4.4 Dretske on Knowledge of One's Knowledge and Knowledge of Natural Laws

A problem with Dretske's analysis of knowledge is its reliance on the concepts of causality and causal support, which many think themselves cry out for a satisfactory analysis. This however can be regarded as a minor drawback, because nothing relevant appears to hinge on the particular account of causality and causal sustain one may wish to endorse.

More relevant is the fact that unlike Nozick's tracking analysis, Dretske's information-theoretic analysis of knowing does not provide a wholly general account of factual knowledge. Dretske cheerfully acknowledges that his account cannot encompass necessary truths: as there are no possible alternatives to them, necessary truth have an informational measure of zero. This means, Dretske hastens to add, that, if natural laws are conceived to be nomically necessary, his account will not be able to handle them either (Dretske 1981a, 264 f.); his analysis is thus admittedly restricted to *perceptual* knowledge.

In my opinion, Dretske's assessment of the scope of his own account of knowing is far too cautious. In this section I shall first consider how it could be extended to knowledge of one's knowledge; and then I shall argue that the possibility of devising an information-theoretic account of knowledge of 'general' facts is not ruled out by the fact that natural laws are (in some sense) necessary.

Let me start from knowledge of one's knowledge. The information-theoretic definition of knowledge has of course the consequence that knowing that one knows that p is not *analytically* entailed by (perceptually) knowing that p. But Dretske (1991, 191) also suggests that 'normally we will not (perhaps cannot) know *that* we know or *when* we know'. How are we to take this statement? Is Dretske espousing Williams' scepticism

about our possibility of knowing that we know? What does 'normally' mean here? That S knows that *p* certainly cannot be known (by S) through perception, let alone introspection (this is because of the externalist character of Dretske's analysis; contrast Prichard's passage quoted in section 3.1 above). However, Dretske's theory does not seem to exclude the possibility that it could be known in some other way. What concept of (higher-order) knowledge is he using then? Even though, as far as I know, Dretske has never explicitly committed himself on this point, his definition of perceptual knowledge does suggest — it seems to me — a natural generalization which would throw some light on *when* (if ever) it could be appropriate to say that S knows that she knows that *p*.

On the basis of Dretske's approach, it seems natural to say that S knows that she knows that *p* just in case (i) she knows that *p*, and (ii) her belief that she knows that *p* is caused (or causally sustained) by the information that the communication channel through which she received the information that *p* is reliable. Although the existence of various laws of nature will contribute to the communication channel's reliability, there is no reason to suppose that the fact that it *is* reliable should itself be nomically necessary. As long as the communication channel is itself regarded as a source of information, there will be alternative possibilities to its being reliable. Therefore the reliability of the communication channel will surely generate information.

Again, one should not require that the information that the communication channel is reliable exclude all alternative possibilities, but only those possibilities which are actually relevant. If our voltmeter was thoroughly checked and calibrated two hours ago (e.g., by measuring the independently known values of the voltage drop across different points of a familiar electric circuit), one will not need to exclude the (irrelevant) possibility that the leads have meanwhile unpredictably changed their resistance, in order to receive the information that the voltmeter *is* reliable. And if one does receive the information that the voltmeter is reliable, why should not one *know* that one knows that the voltage drop across the resistor in the (unfamiliar) electric circuit is 7 volts? So it seems to me that Dretske's account of perceptual knowledge does provide a framework within which one can raise the issue of *when* one knows that one knows that *p* with some prospects of success.

So much for knowledge of one's knowledge. Let me now address the issue of our knowledge of natural laws. The fact that Dretske's analysis of perceptual knowledge

does not account for our knowledge of natural laws is not, in itself, a defect. For it is obvious that the truth of statements expressing natural laws is somehow *inferred* from perceptual knowledge, but it is not itself an object of perceptual knowledge. However, it seems to me that Dretske's information-theoretic approach would face serious difficulties if it did entail the claim that nomically necessary truths (laws of nature) do not generate information. If this were the case, not only would we be prevented from perceptually knowing that a given law-like statement is true, but we could not even *test* true law-like statements against experience, because there would be no set of possible alternatives to choose from in the first place. Suppose, for example, that it was a natural law that rays of light are propagated curvilinearly in gravitational fields; if this fact's being a natural law meant that it could not generate information, then the law-like statement expressing it could not be tested by, say, astronomical observations, because there would be no possible alternatives the observation of phenomena consistent with the law could be seen as eliminating. We would then be compelled to say that a large number of law-like statements are not amenable to empirical testing. If they are true they are necessarily so and hence do not admit of any possible alternative, whereas if they contradict a true law-like statement they are necessarily false and hence constitute no possible alternative to anything. If the necessity of natural laws entailed that they could not generate information, the only law-like statements amenable to empirical testing would be in fact those false law-like statements (if any) which do not contradict any statement expressing a (true) natural law.

However, Dretske's argument that natural laws cannot generate information because they are nomically necessary seems to me hardly compelling, because, as I shall presently argue, the alternatives a state of affairs has to allow in order to generate information need only be *epistemically*, not metaphysically possible.

One way to circumvent the claim that natural laws cannot generate information because they are nomically necessary is to appeal to Kripke's (1972) distinction between the *metaphysical* categorization of truths into necessary and contingent, and the *epistemological* categorization of truths into a priori and a posteriori. If Kripke's claim that certain kinds of identities (e.g., that Hesperus = Phosphorus, or that heat = the motion of molecules) that can be known only a posteriori are metaphysically necessary is correct, then there appear to be many states of affairs which generate information even if they are (metaphysically) necessary. Of course Kripke's point is not made in

information-theoretic terms, but his account of how the obtaining of certain states of affairs may be known a posteriori suggests that necessary states of affairs *must* be capable of generating information and causing (or causally sustaining) people's beliefs. And even if it is highly doubtful if natural laws can be properly described as being *general* states of affairs, the fact that they are, in some sense, nomically necessary need not prevent them from generating information and causing (or causally sustaining) people's beliefs.

Dretske himself seems to accept Kripke's distinction between what is necessary and what is knowable a priori (see Dretske 1981a, 222). But if one does not like the sort of use rigid designators and natural kinds are put to in Kripke's approach, a similar point can be made without assuming the existence of necessary truths that are knowable only a posteriori. All one needs to assume is the existence of necessary truths that can be known in the traditional, a priori way. We saw that Dretske's definition of '*K* knows that *s* is *F*' contains a reference to *K*'s background knowledge: if one already knows that *s* is either red or blue, a signal that eliminates the possibility of *s*'s being blue will be sufficient to support one's knowledge that *s* is red. Apparently, Dretske construes *K*'s background knowledge as being itself an instance of *perceptual* (a posteriori) knowledge. But there is no reason why this should be necessarily so. It is surely reasonable to suppose that knowledge of logical and mathematical truths could also contribute, in some circumstances, to the elimination of certain possibilities about the source.

Let us consider the following example. Ann's teacher has just drawn a triangle on the blackboard. Ann knows how to use a measuring rod, and she also knows that the figure drawn by her teacher is a triangle. But she does not know the geometrical truth that the sum of the lengths of any two sides of a triangle always exceeds the length of the third. However, her ignorance of this necessary truth need not prevent her from getting to know, by carrying out the appropriate measurements and making a simple arithmetic calculation, that *this* triangle is such that the sum of the lengths of AB and BC exceeds the length of AC. My claim is that if Ann gets to know that much, she will have acquired a posteriori knowledge of a (geometrically) necessary truth.

By saying that the proposition, 'the triangle drawn by Ann's teacher is such that the sum of the lengths of AB and BC exceeds the length of AC' is a necessary truth I mean that it follows deductively from a theorem of geometry under the supposition that there is such a triangle. Of course if that proposition is analyzed in terms of Russell's

theory of descriptions, it will come out as merely contingently true. I would rather analyze it in terms of Strawson's theory and say that it is a necessary truth whose utterance 'presupposes' a commitment to the existence of the triangle-token in question. This is, however, a minor issue. The point is that it is not the existence of the triangle-token, but the relationship between the lengths of its sides that Ann can get to know through perception. That the obtaining of this relationship in the triangle in question is (geometrically) necessary is what really matters for my argument. And I am inclined to believe that the obtaining of this relationship in the triangle in question is indeed necessary because if one were to deny its necessity on the ground that Ann's teacher could have failed to draw the triangle on the blackboard, one would then be committed to deny that any contingent individual is necessarily identical to itself, which seems to me a rather undesirable consequence.

But even if it is not granted that the relevant relationship between the sides of the triangle drawn by Ann's teacher obtains necessarily, the fact that Ann can get to know through perception of the obtaining of a state of affairs which admits no relevant alternatives appears to follow from the supposition that Ann *knows* that the figure drawn by her teacher is a triangle. This background knowledge of hers crucially restricts the range of possible alternatives she must be able to exclude in order to learn that the figure drawn by her teacher is such that the sum of the lengths of AB and BC exceeds the length of AC. Since Ann *knows* that the figure drawn by her teacher is a triangle, the possibility that the relationship in question may not obtain because the figure is not a triangle represents no genuine alternative to the actual state of affairs. Therefore we are left once more with the fact that Ann comes to know through perception of the obtaining of a state of affairs which, given her background knowledge, admits no possible alternatives.

Now, there is apparently no reason why Dretske's account of knowing should not apply to Ann's knowledge of the relationship between the lengths of the sides of the triangle drawn by her teacher. True, if her geometry were good enough, she could know a priori that this relationship obtained, and indeed that it was necessary. But this does not mean that she cannot know a posteriori, by means of her measuring rod, that this relationship obtains in the triangle-token in question (though of course without knowing that it obtains necessarily). From her standpoint, the hypothesis that the sum of the lengths of AB and BC may be less or equal to the length of AC is indeed a relevant

alternative! The fact that deeper geometrical knowledge on her part would stop it being so does not mean that it is not so in her present epistemic situation. Therefore I conclude that in certain circumstances *lack* of logical and mathematical knowledge may allow a knowing subject to gain a posteriori knowledge of necessary truths. And this again provides a reason for thinking that the necessary character of natural laws need not prevent them from generating information and causing (or causally sustaining) people's beliefs.

A third reason for thinking that natural laws may indeed generate information is of course that, even if they are commonly thought to be 'nomically' necessary, they are certainly not *logically* necessary. Let us suppose that Newton's inverse square law was nomically necessary; this would not commit us to claim that its negation was logically contradictory. There are, as it were, many logically possible 'worlds' in which the prevailing laws of nature are different to ours. Or, to use a theological metaphor, God could have created a world in which Newton's inverse square law did not obtain. Even though the diference between logical and nomical necessity has sometimes been denied, its *prima facie* plausibility does support the thesis that natural laws could be, in some sense, necessary and yet generate information.

The arguments I have given so far do not question Dretske's (1981a, 264 f.) assumption that natural laws are (nomically) necessary. But are they really so? Dretske himself once argued that they are not. According to Dretske (1977), they are 'contingent relationship between properties' expressed by law-like statements of the form:

(*) *F*-ness → *G*-ness,

where the connective '→' can be read as 'yields' in the case of simple qualitative laws, while it can stand for various matemathical relations in the case of laws involving quantitative expressions. Let us suppose that (*) expresses a qualitative law. If (*) is true, it seems plausible to infer that this particular *F must* be *G* (even if Dretske acknowledges that he cannot provide a conclusive argument to this effect). However, (*) itself will be merely contingently true:

> Although true statements of the form of [(*)] are not themselves *necessary* truths, nor do they describe a modal relationship between the respective qualities, the

> contingent relationship between properties that is described imposes a modal quality on the particular events falling within its scope
>
> (Dretske 1977, 263-264)

On this view of natural laws, if there are any truths the epistemic accessibility of which may be thought to be jeopardized, these are truths of the form, 'This *F* is (necessarily) *G*'. I bracketed the modality because the problem is not, in the first place, whether we may come to know that *this F is necessarily G*, but merely whether we may come to know that *this F is G*, if it is necessarily so. The only reason why the epistemic accessibility of contingent truths like (*) may also seem to be jeopardized by this view of natural laws is that epistemic access to contingent truths like (*) is usually taken to depend on epistemic access to the events falling within their scope. But there is no doubt that the modality of natural laws cannot constitute, on Dretske's (1977) account, an obstacle to their generating information.

As for the problem of the epistemic accessibility of truths of the form 'This *F* is *G*', it seems to me that every theory preventing us from having *perceptual* knowledge of truths of the form 'This *F* is *G*' (where this *F* is *necessarily G*) should be rejected simply because it is an hard fact that we do have *perceptual* knowledge of truths of that form. This hard fact could be denied only by supposing that all the truths of the form 'This *F* is *G*' of which we can have perceptual knowledge are such that the *F* is only contingently *G*. But this would mean supposing that all natural laws relate unobservable properties, i.e., that the relations between the macroscopic qualities of things are never a matter of necessity, which surely is far from uncontroversial.

Moreover, even if this assumption were true, it would still be only contingently true. Even if God the Father did write down all the laws of physics and the initial distribution of matter and energy, and then left to the Logos the job of calculating all the macroscopic properties and regularities entailed by those laws and initial conditions, He could as well have written down each and every law that He wanted the universe to *display*, and then left to the Logos the task of so arranging the initial conditions as to secure their mutual consistency (Cartwright 1994, 291, theologically revised). But then, if it is only contingently true (if it is true at all) that all natural laws relate unobservable properties and that no relation between the observable qualities of things obtains necessarily, the distinction between those *F*'s that can be perceptually known to be *G*'s and those *F*'s that cannot be perceptually known to be *G*'s appears totally arbitrary. If

God had chosen to carry out creation according to the second option which has just been described, some *F*'s would be observably *G*'s and yet incapable of being perceptually known as such. This seems to me totally absurd — even more absurd, in fact, than supposing that the relationship between the lengths of the sides in the triangle drawn by the teacher should be inaccessible to Ann's measurements (plus calculations) because of its being geometrically necessary.

The conclusion I want to draw from these remarks is not that Dretske's analysis, failing to account for many actual or possible instances of perceptual knowledge, has to be rejected. What has to be rejected is rather Dretske's assumption that necessary states of affairs cannot generate information. I have tried to substantiate the claim that they can, because the alternatives a state of affairs has to allow in order to generate information need only be *epistemically*, not metaphysically possible. If so much is granted, then the modality of natural laws (or of the events falling within their scope) need no longer be seen as an obstacle to their generating information. And there is no reason why the information that a law-like statement is *false* could not be 'nested' in some experimental results. For experiments could be viewed, at least in connection with theoretical science, as a way of getting the information that certain relevant alternatives do not obtain.

Still, well-known difficulties about the justification of inductive practices and especially the Duhemian point that the empirical sciences cannot perform genuine *experimenta crucis* may be thought to entail that no belief in a natural law can be caused (or causally sustained) by the information that (necessarily) all *F*'s are *G*'s — or that *F*-ness yields *G*-ness. If our theoretical choices are hopelessly underdetermined by the finite evidence provided by perception, it may seem as if our beliefs about natural laws will always admit of (relevant) alternatives that cannot be conclusively eliminated. But even if (inferential) knowledge of natural laws is admittedly a much trickier matter than (perceptual) knowledge of particular states of affairs, standard difficulties about the justification of inductive practices and the above-mentioned Duhemian thesis do not entail that an information-theoretic approach to (inferential) knowledge of natural laws must have as a result that such knowledge is impossible. For if epistemic justification fails to be a necessary condition of knowledge, a satisfactory analysis of (inferential) knowledge of natural laws can be wholly independent from the issue of the justification of our inductive practices. Indeed, for an information-theoretic approach to factual

knowledge the issue is not whether we can be *justified* in believing that (necessarily) all *F*'s are *G*'s, but whether the inductive practices involved in our belief that (necessarily) all *F*'s are *G*'s can yield the *information* that (necessarily) all *F*'s are *G*'s. But then, it seems reasonable to suppose that the application of certain inductive practices to specific empirical problems may yield the information that (necessarily) all *F*'s are *G*'s even without enabling us to devise a proof that the evidence to which we have access entails that (necessarily) all *F*'s are *G*'s, just as it seems reasonable to suppose that the operation of our perceptual apparatus may yield the information that this *F* is *G* even without enabling us to devise a proof that (say) our sense data entail that this *F* is *G*. Thus I conclude that the generalization of Dretske's information-theoretic analysis of perceptual knowledge to more 'theoretical' kinds of empirical knowledge presents no insuperable difficulty.

4.5 Is There a 'Social' or 'Pragmatic' Aspect to Knowledge?

There is a further problem with Dretske's analysis of knowledge which I want to address in this chapter. This problem stems from a certain ambiguity in his treatment of what he calls (Dretske 1981b; also 1981a, 132-134) the 'social' or 'pragmatic' dimension of knowledge. A discussion of this ambiguity will provide us with the extra benefit of a better understanding of the point of the 'relevant alternatives' answer to the sceptic.

Let me start with one of Dretske's own examples (see Dretske 1981b, 375). The fuel gauge mounted on my car may be thought of as a reliable means for knowing that I still have enough petrol when driving in the city. But would the same gauge be good enough for knowing that there is sufficient liquid coolant surrounding the reactor on Three Mile Island? As the stakes go up, it may look as if we need to exclude more alternatives (e.g., a malfunction in the instrument) in order to be able to *know* what the gauge is supposed to tell us. Whether S may be said to know that *p* may thus seem to depend, in certain circumstances, on pragmatic factors other than S's evidential status vis-à-vis the state of affairs that *p*. It may seem to depend, in fact, on the *relevance* of what is known for the knower's own interests and purposes.

Now, there are passages in which Dretske seems to suggest that this is precisely how things are. For example, he writes that our decisions about the reliability of the

communication channel (i.e., our decisions about which alternatives are to be viewed as irrelevant for the sake of the flow of information) are 'responsive to the interests, purposes, and, yes, values of those with a stake in the communication process' (1981a, 133). This sort of remark seems to imply that what is taken to be a reliable channel for some purposes may not be counted as a reliable channel for different purposes, and that although the fuel gauge mounted on my car can provide information about the state of my tank, it would provide no information about the liquid coolant surrounding the reactor on Three Mile Island.

However, Dretske also writes that although many examples seem to indicate that whether S knows that *p* may depend on pragmatic factors other than S's evidential status vis-à-vis what is known,

> most of these examples show nothing of the kind. These factors affect, not *whether* something is known, but whether it is reasonable to *say* you know or to *think* you know.
>
> (Dretske 1981b, 367)

This passage seems to deny that S's knowing that *p*, as opposed to S's claiming or believing that she knows that *p*, is affected in any relevant way by S's interests, purposes and values. And in fact Dretske's assessment of the fuel gauge case turns out to be that there is no reason

> why a standard automobile gauge, transplanted from the automobile to the nuclear power plant, functioning as the *only* indicator of coolant level, should not, assuming it continues to function reliably (as reliably as it did in the automobile), be able to do precisely what the more expensive instruments do [...]. I admit that the operators *should not* rely on a single gauge, and certainly not one manufactured under such casual quality control, but if they *do* rely on it, I don't see any basis for denying that they know.
>
> (Dretske 1981b, 376)

Now, the reason of the apparent inconsistency in Dretske's treatment of the pragmatic dimension of knowledge is, as Cohen (1991, 21-23) points out, his failure to appreciate that there are two different kinds of contextual relativity, which he runs together under the heading, 'social or pragmatic aspects of knowledge'. Knowledge may be thought to be context-relative in the sense that whether S knows that *p* depends on certain extra-evidential features of S's circumstances — this is what the pain-simulators and the bent

stick examples are meant to emphasize. We may call this phenomenon the 'extra-evidential context-relativity of knowledge'. But knowledge may also be thought to be context-relative in the sense that whether S knows that *p* depends on the intentions, purposes, etc. of the knowledge-attributor. We may call this different phenomenon the 'context-of-attribution relativity of knowledge'.

Now, Dretske's failure to appreciate the distinction between these two different sorts of context-relativity has two undesired effects. First, it makes it appear as if his epistemology is committed to the independent and unnecessary claim that there is indeed a phenomenon like the context-of-attribution relativity of knowledge. Secondly, it creates the impression that Dretske may be suggesting that the interests, purposes and values of people contribute to determine which extra-evidential factors are relevant in deciding whether S knows that *p* even in such cases of purely extra-evidential context-relativity as the pain-simulators and the bent stick examples. But in fact Dretske's account of knowing does not support, as it stands, either of these claims.

Dretske's account entails that whether S knows that *p* depends on which alternative possibilities to *p*'s being true are *relevant* possibilities, that is, possibilities which S's evidence must be able to exclude. Since Dretske's account does not rely on possible world similarities to decide which alternative possibilities are relevant possibilities, if we buy his analysis we will have to choose one of the following options. We will have to say that which alternative possibilities are relevant possibilities depends either (i) on the extra-evidential features of S's circumstances, or (ii) on the interests, purposes and values of the knowledge-attributor[8], or (iii) on both factors.

Challenged by Cohen to clarify his position, Dretske (1991, 191-196) admits to having been confused and definitely opts for (i):

---

[8] Hookway (1990, 201-203) maintains that which alternative possibilities are *relevant* possibilities for the purpose of ascribing knowledge depends on the 'background knowledge of the person who makes the ascription': somebody who is unable to tell Judy from her identical twin Trudy and yet claims to be seeing Judy across the street can be warrantedly said to *know* that it is Judy that he is seeing provided the knowledge-attributor is confident that Trudy is away on holiday. Here I cannot go into Hookway's motivation for adopting this view, which has to do with the role of knowledge-attributions in learning from testimony. However, if I understand his proposal correctly, his suggestion is not merely that one knowledge-attributor may be warranted in saying that S knows that *p* while another knowledge-attributor may be warranted in saying that she doesn't, but that *whether S knows that p* depends upon the background knowledge of the attributor. This involves a pretty radical relativization of knowledge, because saying that knowledge-possession is relative to the backgorund knowledge of the attributor, unlike saying that it is relative to the knowledge-attributor's interests, purposes and values, makes it impossible to account for the fact that S can be simultaneously described as knowing and as failing to know that *p* by appealing to an ambiguity in the concept of knowledge employed in the two conflicting descriptions.

> What *is* relative to the attributor and the circumstances of utterance (the context of utterance, if you will) is what the knower is *said* to know [...]. The fact that the circumstances of the speaker can affect what the speaker is *saying* someone (himself or another) knows [...] is *not* relevant to understanding the concept of knowledge itself — what it takes to know what we are said to know.
>
> (Dretske 1991, 192 f.)

This statement sanctions his (1981b) assessment of the fuel gauge case and retracts his (1981a) claim that knowledge is responsive to the interests, purposes, and values of those with a stake in the communication process.

I am perfectly happy to go along with Dretske's option for (i). It has been argued, though, that a pure and simple denial of the context-of-attribution relativity of knowledge will make it impossible for someone upholding the 'relevant alternatives' answer to the sceptic to account for the persisting appeal of sceptical arguments (see Cohen 1991, 27-28). If sceptical alternatives are *objectively* irrelevant because it makes no sense to suppose that a shift in one's cognitive standards could turn them into *relevant* alternatives, then it is not clear how the sceptic could ever convince us to take them seriously. The persistent appeal of sceptical arguments apparently comes from the fact that they can induce us to revise our usual standards of relevance by taking into account alternative possibilities that we would normally dismiss as *irrelevant*. Holding 'the circumstances of S fixed, one speaker may say "S knows *p*" while another says "S does not know *p*" without it being the case that they have contradicted each other' (Cohen 1991, 23) precisely because which alternative possibilities are relevant seems to depend on our interests, purposes and values. It is because of this situation that the sceptic can induce us to give up our knowledge claims by convincing us to endorse more liberal standards of relevance (that is, stricter standards for knowledge). If knowledge were not to display such context-of-attribution relativity, the persistent appeal of sceptical arguments would be much more difficult to be accounted for.

I think that this argument is far from compelling, and that an externalist epistemologist espousing a 'relevant alternatives' answer to the sceptic can provide a perfectly adequate account of the persistent appeal of sceptical arguments without assuming the context-of-attribution relativity of knowledge. Among the factors involved in deciding whether S knows that *p*, one should take into account the fact that when the attributor is the same person as the knower, her interests, purposes and values can affect whether S knows that *p* not by affecting the range of the relevant alternative possibilities

to *p*'s being true, but by weakening S's disposition to believe that *p* on the basis of her information. In other words, when the attributor is the same person as the knower, whether S knows that *p* does depend on the interests, purposes and values of the attributor, but only because these can affect (that is, *causally* affect) S's disposition to *say* or *claim* that she knows that *p*, and hence her disposition to *believe* that *p*. If the sceptic allures me into thinking that my information must suffice to exclude a whole class of (objectively) irrelevant alternative possibilities if I am to know that *p*, I can end up mistrusting my information and suspending my judgement on *p*. But this is not so because whether I can legitimately be said to know that *p* is itself relative to the context-of-attribution, but because whether I can legitimately be said to know that *p* depends on whether I actually *believe* that *p*. And the sceptic can induce me to regard my belief that *p* as epistemically suspect by persuading me to attach an undue relevance to what are in fact irrelevant alternatives.

This point highlights an interesting difference in the way sceptical arguments affect internalist and externalist — especially naturalistic — approaches to knowledge. The aim of the sceptic's strategy of imposing exceedingly strong requirements on knowledge is clearly that of robbing us of most of the knowledge we usually believe we have. However, this goal is accomplished in different ways depending on the sort of knowledge one is supposed to have. If our knowledge is conceived in an internalist fashion, the sceptical arguments will be meant to show that most of our beliefs do not satisfy the requirements that the sceptic takes to be constitutive of the very idea of knowing. The sceptic will harness the internalist conviction that each and any belief which is to qualify as an instance of knowledge must be epistemically transparent to the knowing subject to prove that no relevant belief about the external world can satisfy the necessary conditions for knowledge. But if knowledge is an external, factual connection of the knower to the world, the only way in which the sceptic will be able to rob us of our knowledge will be by weakening our disposition to believe what we believe by inducing us to regard all our information as ambiguous. If knowledge is an external relationship to the world, the sceptic's arguments will have, as it were, a rhetoric rather than an argumentative force. By this I mean that their success will depend on how effectively they can induce us to give up our usual concept of knowledge (according to which being able to prove beyond any possible doubt that one knows that *p* is not a necessary condition for knowing that *p*) and to endorse the sceptic's cunningly contrived

alternative, rather than on any intrinsic feature of knowledge. In other words, while internalist knowledge is bound to be doubtful until the sceptic's arguments have been rejected (this was the insight that motivated Descartes' project of 'pure inquiry'), externalist knowledge is 'out there' all right even if sceptical doubts are yet to be voiced. And out there it will remain, unless we are allured by the sceptic's arguments into distrusting that very knowledge that we already have.

Now, to go back to Cohen's proposal, it seems to me that beyond showing that it is possible to acknowledge the role played by the interests, purposes and values of the knower in accounting for the persistent appeal of sceptical arguments without endorsing the view that knowledge is relative to the context-of-attribution, there is not much to say against Cohen's view, except perhaps that it yields a proliferation of senses of knowing which does not seem to correspond to any actual linguistic practice. The upshot of Cohen's proposal is in fact that knowledge is an ambiguous concept, which has different meanings in everyday discourse and in the sceptic's mouth, whereas I think that one should stick to the claim that the sceptic is simply *wrong* in denying that in many ordinary circumstances we know the things we believe we know. I do not think that the sceptic's concept of knowledge should be granted the same consideration as the standard, everyday concept of knowledge: the sceptic's concept of knowledge simply is *not* a relevant alternative to the standard one!

This may seem a merely verbal point. But while we need our ordinary concept of knowledge to explain (or predict) the success of a wide range of human activities — I am thinking, for example, of the attribution of knowledge as a way of explaining a subject's ability to successfully perform certain operations, of the appeal to a witness' knowledge of the facts as a way of learning things one is not already acquainted with, and of the reliance on the expertise of knowledgeable individuals for the purpose of carrying out certain tasks — the sceptic's concept of knowledge is useless outside epistemology classes. This is why I would rather say that the concept of knowledge is *vague* than that there are many different senses of knowledge around: from the car-driver's through the nuclear engineer's to the sceptic's.

If the concept of knowledge is somewhat vague, the extra-evidential features of S's circumstances will in many cases determine whether S knows that p, but there will be border-line cases in which no clear-cut answer to the question is forthcoming, because the extra-evidential features of S's circumstances do not unambiguously determine

whether a given possibility *is* a relevant alternative to p. Think again of the pain-simulators example: would one know that N is in pain if the custom of pain pretence were not absolutely general among his fellow-citizens, but concentrated in some regions of the country? Maybe he does not live in one of these regions. But he may live very close to one of these regions. Or he may be the first person we meet after we have visited these regions. Some ingenuity will suffice to devise cases in which the extra-evidential features of the knower's circumstances will not provide any clear-cut answer to the question, 'Does S know from N's behaviour that N is in pain?'[9].

Several conditions have been proposed as an account for our intuitions about 'relevant' alternatives. If the concept of knowledge is indeed vague, then it is an adequacy requirement on such conditions that they exhibit a comparable degree of vagueness. But it is not really important for my argument which account best captures our intuitions about 'relevant' alternatives, as long as these intuitions are agreed to be insensitive to the context of knowledge-attribution.

I think that the vagueness of the concept of knowledge should not be regarded as disappointing: many concepts are vague and yet perfectly meaningful. And more often than not, vague concepts are useful precisely because of their vagueness: no advantage would be gained by replacing them with more precise concepts. Moreover, in (a realist) epistemology it is not the concept of knowledge that does the normative work, but the concept of truth: we want our methodological rules to be truth-conducive rather than knowledge-conducive (the fact that truth does not represent our only cognitive goal is irrelevant to the present issue). No doubt, if S can legitimately be described as knowing that p, p itself will have to be true. But p can be definitely true even if the question whether S can legitimately be described as knowing that p has no clear-cut answer because the extra-evidential features of S's circumstances do not determine a definite set of relevant alternative possibilities to p's truth.

---

[9] Another good example (the Gadwall duck and the Siberian Grebes) is discussed in Dretske (1981b, 368-370).

Chapter 5

The Need for Epistemic Justification

We saw in the second chapter that the point of most epistemic theories of truth is, ultimately, to increase our chances of cognitive success, which are alleged to be negligible if truth is intended as correspondence to a knowledge-independent reality. For how could we hope to gain any knowledge of the world if it is a consequence of our notion of truth that the evidence to which we have access will forever underdetermine the truth of even the most cautious among our beliefs? Confronted with this question, one can either resolve to close the gap between evidence and truth by brute force by defining the latter as a function of the former, or — holding on to an evidence-transcendent notion of truth — try to substantiate the claim that having knowledge does not entail being able to rule out *all* possible alternatives to the actual state of the world. While the former option falls squarely into the internalist tradition which has kept up the Cartesian legacy within the fallibilist context of our century, the latter option has yielded various sorts of externalist accounts of knowledge, whose emphasis is more on the actual relationship of the knowing subject to the world than on the self-transparency of the knowing subject's epistemic situation. In particular, the externalist accounts of knowledge discussed in the last chapter are *naturalistic* accounts according to which the actual relationship of the knowing subject to the world is all-important, and S's knowing that p does not necessarily involve S's being justified in believing that p.

In this chapter I propose to take up the internalist challenge to naturalistic accounts of knowing as expressed by the hologram and clairvoyance counterexamples presented in section 4.2. I want to assess whether naturalistic accounts of knowing ought

really to be seen, as Bonjour (1980, 70) puts it, 'as simply abandoning the traditional idea of epistemic justification or rationality and along with it anything resembling the traditional conception of knowledge'. Addressing this issue will engage me in a wider discussion of the role and character of epistemic justification in our argumentative practices.

Since Nozick's tracking analysis fails to provide an adequate account of our pre-philosophical intuitions about knowledge, the account of knowing I tentatively endorse is Dretske's information-theoretic analysis as construed in sections 4.3 to 4.5 above. I say 'tentatively' because I am aware that the lack of a fully worked-out account of our intuitions about 'relevant' alternatives may cast some doubts about its overall adequacy. At any rate, the specific features of Dretske's account are not really decisive for the argument of this chapter, because the internalist challenge to naturalistic accounts of knowing is perfectly general and affects Dretske's analysis only in so far as it denies that (internal) epistemic justification is a necessary condition for knowledge. Accordingly, the argument of this chapter will also be so general as not to depend on the distinctive features, let alone the correctness, of Dretske's proposal.

Section 5.1 describes the 'epistemization' conception of justification which underlies the main internalist objection to naturalistic accounts of knowing. Section 5.2 introduces the distinction between knowledge-attributions and knowledge-claims and argues that the hologram and clairvoyance examples cannot be taken to express unambiguously internalist intuitions about knowledge. Section 5.3 rehearses the argument from the existence of 'animal' knowledge to the superfluity of internal justification for knowledge. This argument is meant to support the claim that a thoroughly naturalistic approach to knowledge is preferable to those externalist approaches that, while denying that all genuine instances of knowledge must be epistemically transparent to the subject to which they belong, nevertheless continue to regard some kind of internal epistemic justification as a necessary condition for knowledge. Sections 5.4 and 5.5 explain why the existence of *internally* justified beliefs, though unnecessary for knowledge, is necessary for our argumentative practices and for the formulation of cognitive prescriptions (this is the insight that must be credited to the internalist approach to epistemic justification). Section 5.6 emphasizes what I call the 'pragmatic' character of epistemic justification, and section 5.7 highlights the 'non-psychologistic' features of the picture of knowledge and epistemic justification I am trying to delineate.

5.1 Justification and the 'Epistemization' of Beliefs

Beliefs may be true by accident, but of course a belief which merely happens to be true will not qualify as knowledge. Within the framework of the Justified True Belief account of knowing, it was epistemic justification that turned true beliefs into knowledge: epistemic justification was said to 'epistemize' true beliefs. It provided the knower with a 'grip' on the truth of the believed proposition, making of knowledge — as opposed to merely true belief — something of which the knower was, as it were, in control.

In the last two chapters we came across various externalist requirements (reliability, tracking, etc.) which have been suggested — after the collapse of the JTB account of knowing — as conditions which true beliefs ought to satisfy in order to qualify as knowledge. Such requirements are 'externalist' because the cognitive subject need not be aware of their being satisfied for a true belief which satisfies them to qualify as knowledge. On the other hand, the requirement of epistemic justification, at least as it was commonly interpreted in the framework of the JTB account of knowledge, had to be 'internalist', if it was to guarantee the internalist character of knowledge.

Nowadays, nobody believes in the adequacy of the JTB account of knowing, but a number of epistemologists are still convinced that epistemic justification is a necessary condition for knowledge. They believe, roughly, that no one can legitimately be said to know that *p* who cannot provide any (compelling) reason why *p* should be regarded as true rather than false. This is the attitude which underlies such alleged counterexamples to naturalistic analyses of knowing as were described in the last part of section 4.2 above. Why cannot the person who sees the vase-activated hologram be said to *know* the true proposition that there is a vase before her? Because the reason why she believes that proposition is that she falsely believes that she can *see* the vase, whereas what she actually sees is but a vase-hologram. Why cannot clairvoyant Norman be said to *know* the true proposition that the President is in New York City? Because apart from his belief, he cannot provide any evidence at all for the proposition that the President is in New York City.

Such examples are meant to capture the internalist idea that no subject who is altogether ignorant of her own epistemic situation can reasonably be said to *know* what she truly believes, no matter how reliable her belief-forming processes may be. Lehrer (1990, 162) expresses this feeling by saying that all *purely* externalist theories of

knowledge — that is to say, all those theories according to which internal epistemic justification is not even a necessary condition for knowledge — 'share a common defect, to wit, that they provide accounts of the possession of information rather than of the attainment of knowledge'. In order to show that the case provided by the hologram and clairvoyance examples does not suffice to refute a naturalistic (i.e., in Lehrer's words, a purely externalist) approach to knowledge, I shall outline the main features of an internalist and pragmatic conception of epistemic justification which provides an account of those examples according to which their protagonists possess genuine knowledge of the true propositions they are described as believing. The purpose of this chapter is not, however, to work out a detailed theory of (internal) epistemic justification, let alone an exhaustive set of criteria of epistemic justifiedness. My aim is merely to offer some arguments in support of an internalist and pragmatic theory of epistemic justification, in order to show that the conjunction of a naturalistic account of knowing with an internalist and pragmatic account of epistemic justification can provide a suitable framework for the project of epistemology.

5.2 Knowledge Attributions and Knowledge Claims

The hologram and clairvoyance examples are meant to show that epistemic justification must be internalist. This, I think, they successfully do, provided a belief's being epistemically justified is not *ipso facto* taken as evidence of its being likely to be true. After the decline of epistemological dogmatism, the fact that epistemic justification must be internalist represents perhaps the strongest motivation for the claim that *knowledge* must be internalist. This is the second half of the story recounted in section 3.3: since there seem to exist good independent reasons for an internalist view of epistemic justification, it would be unfair to see present-day knowledge-internalism as a *mere* watering-down of epistemogloical dogmatism. However, I shall argue that it is false that the fact that epistemic justification must be internalist entails that *knowledge* must be internalist.

Some writers tend to conflate these two issues. Having told his story about Norman's clairvoyance power, Bonjour goes on to ask: '*why* should the mere fact that such an external relation obtains mean that Norman's belief is epistemically justified,

when the relation in question is entirely outside his ken?' (Bonjour 1980, 63). Bonjour's implication that the reliability of Norman's clairvoyant power does not *justify* his belief that the President is in New York City is perfectly acceptable. But Bonjour takes this to entail that Norman cannot be described as *knowing* that the President is in New York City either: 'From his standpoint, there is apparently no way in which he *could* know the President's whereabouts' (Bonjour 1980, 62). Now, one might think that the reason why Bonjour regards his story as a refutation of the claim that Norman could *know* that the President is in New York City is merely that he believes that epistemic justification is a necessary condition for knowledge. But it is revealing that the externalist account of epistemic justification that Bonjour's paper assumes as paradigmatic is not taken from an analysis of justification, but from Armstrong's (1973) reliabilist analysis of *knowledge*. And while Bonjour appears to be perfectly aware that Armstrong's analysis is formulated 'in terms of knowledge rather than justification', his insistence on construing Armstrong's reliabilism as entailing 'that beliefs satisfying his [Armstrong's] externalist criterion are epistemically justified or rational' (Bonjour 1980, 57) betrays the conviction that an externalist analysis of knowledge cannot but arise by replacing an internalist view of justification with an externalist one. Therefore it would seem that the reason why Bonjour regards his story as a refutation of the claim that Norman could *know* the President's whereabouts is more likely to derive from a conflation of the issues of epistemic justification and knowledge than from the quite independent point that epistemic justification is a necessary condition for knowledge.

However this may be, the circumstance that a number of epistemologists have been inclined to believe that the fact that epistemic justification must be internalist entails that *knowledge* must be internalist is likely to be a legacy of their past (or present) commitment to the traditional account of knowledge as justified true belief (see, e.g., Chisholm 1989, ch. 8, esp. 75 f.). If what turns a true belief into knowledge is its being an *epistemically justified* belief, then an externalist account of justification will necessarily yield an externalist account of knowledge. On the other hand, an externalist account of knowledge need not involve an externalist account of justification; yet the influence of the JTB account of knowledge has been so strong that many epistemologists seem unable to understand the attempt to develop an externalist view of knowledge other than as an attempt to provide a new, externalist version of the notion of epistemic

justification[1]. When confronted with an externalist definition of *knowledge*, they will interpret it straightaway as the proposal of an externalist account of *epistemic justification*. They will then proceed to dismiss it as totally irrelevant to 'the analysis of any ordinary concept of knowledge or of epistemic justification' on the ground that it cannot provide any answer to such questions as, 'What can I know?', 'How can I be sure that my beliefs are justified?' and 'How can I improve my present stock of beliefs?' (Chisholm 1989, 76).

This I maintain is a mistake. An externalist view of knowledge will look irrelevant to any traditional epistemological issue only if it is taken to imply commitment to an externalist view of epistemic justification. But such an implication cannot be taken for granted. Indeed, I argued in chapter 3 that it is an *internalist* view of knowledge that implies commitment to an externalist view of epistemic justification. Therefore I propose to show that an accurate blend of externalism in the theory of knowledge and internalism in the theory of epistemic justification may eventually provide the most attractive epistemological option.

Unfortunately, there is a widespread belief that the behaviour of sentences like 'I know that *p*' resembles more closely the behaviour of sentences like 'I have a toothache' than that of sentences like 'I am six feet tall'. Suppose that S is, as a matter of fact, six feet tall. That makes the sentence 'S is six feet tall' true. But does that warrant S's utterance of the claim, 'I am six feet tall'? Of course it doesn't. S can legitimately assert that she is six feet tall only if she has measured her own height, or had it measured by someone relative. Her sentence is made true — loosely speaking: recall the considerations made in section 1.4 above — by the fact that she *is* six feet tall, but her statement is unwarranted unless she can provide some justification for making it. On the other hand, it is often thought that people cannot be wrong about sentences like 'I have a toothache'. The mere fact that I have a toothache is taken to warrant the utterance of the claim, 'I have a toothache'. Epistemologists in the Cartesian tradition have similarly thought that, if S knows that *p*, she must *ipso facto* be warranted to claim

[1] Lehrer (1990, 14), for example, is perfectly aware that externalist philosophers 'may even go so far as to deny that justification is necessary for knowledge'; yet he claims that one may 'do the externalist no injury by looking upon the external connection as providing us with a kind of external justification'. To be sure, proponents of externalist analyses of knowing have sometimes encouraged this construal of their positions by presenting their accounts as externalist analyses of epistemic justification, e.g., Alvin Goldman 1979 (but not Alvin Goldman 1976).

that she does — for how could one know that *p* without being able to tell whether one knows that *p* or merely *apparently* knows that *p*?

I think that the temptation to assimilate the behaviour of sentences like 'I know that *p*' to that of sentences like 'I have a toothache' should be resisted. One might start from the observation that *attributing* knowledge — i.e., describing an individual as possessing some knowledge — is logically distinct from *claiming* knowledge. Even though A's attributing knowledge to B will usually involve A's claiming to know that B has some knowledge, a knowledge-attribution (which can also be a *self*-attribution) can be true without licensing any knowledge-claim at all on the part of the attributee. According to Nozick's or Dretske's view of knowing, the sentence 'S knows that *p*' is made true — loosely speaking — by S's bearing a specific factual connection to the world, but the obtaining of that connection does not warrant, as such, S's claim that she knows that *p*. The fact that I know that *p* makes my utterance of the sentence 'I know that *p*' true, but it does not warrant my claim that I know that *p*, that is, it does not license me to *assert* that *p* or that I know that *p*. Just as the fact that S can be *truly* described as being six feet tall does not suffice to warrant S's claim that she is six feet tall, the fact that S can be truly described as knowing that *p* does not suffice to warrant S's claim that she knows that *p*. In the same way, just as the fact that S can be *justifiedly* described as being six feet tall does not endow S with the capacity of justifying the claim that she is six feet tall, the fact that S can be *justifiedly* described as knowing that *p* does not endow S with the capacity of justifying the claim that she knows that *p*.

As Austin (1946) forcefully argued, knowledge *claims*[2] involve the subject's commitment to provide, on request, an adequate justification for her claim (unless she is joking, acting, or involved in other kinds of activities in which one can utter a descriptive sentence without making a claim). Knowledge *attributions*, on the other hand, albeit involving the *attributor*'s commitment to provide an adequate justification for her attribution, do not require that the *attributee* be able to justify whatever knowledge she is being attributed. For while every knowledge-claim to the effect that *p* involves a self-attribution of the knowledge that *p* (I cannot claim that I know that *p* without attributing to myself the knowledge that *p*), not every knowledge-attribution to the effect that S knows that *p* involves S's self-attribution of the knowledge that *p*. If I attribute to S (S

[2] Knowledge claims are typically performed by uttering sentence-tokens of the form 'I know that *p*'. But a knowledge claim is perhaps 'implicit' in most statements or assertions that *p*.

≠ G.V.) the knowledge that *p*, I attribute to myself the knowledge that S knows that *p* (and am thereby committed to justify the claim that S knows that *p*), but this need not involve S's self-attribution of the knowledge that *p*, and therefore need not involve S's commitment — or ability — to justify the knowledge claim that *p*.

Someone might be inclined to charge my use of the term 'knowledge' with being equivocal on the ground that it introduces an asymmetry between attributions of knowledge to oneself and attributions of knowledge to other people. But I am introducing no asymmetry at all between attributions of knowledge to oneself and attributions of knowledge to other people *qua* knowledge-attributions! *All* knowledge-attributions involve the attributor's commitment to justify the claim that the attributee knows what she is claimed to know. If in the case of *self*-attributions this commitment becomes an *attributee*'s commitment to justify the claim that she knows what she claims to know, this is due to the trivial reason that in the case of self-attributions attributor and attributee are one and the same person.

The distinction between knowledge attributions and knowledge claims provides us with a different standpoint (different, that is, to the traditional internalist perspective) for our assessment of the clairvoyance case. It is perfectly true that Norman cannot *claim* that he knows that the President is in New York City, because he can provide no epistemic justification for the statement that the President is in New York City. But this does not mean that he cannot *know*, as an unsuspecting but completely reliable clairvoyant, that the President is in New York City.

The fact that he cannot claim that the President is in New York City only means that he would not be warranted to make such a claim, it does not mean that he cannot be truly described as *knowing* that the President is in New York City. Should Norman wish to *claim* to have knowledge that the President is in New York City, he would clearly need to assess the reliability of the cognitive method which led him to endorse his belief. As Bonjour points out, he would need to assess the possibility of reliable clairvoyance in general and to decide whether he himself possesses such a cognitive faculty. If he could succeed in carrying out this task and if the outcome of his inquiry were to be favourable, he would then be able to provide a justification for his belief and could be inclined to claim that he not only believes, but knows that the President is in New York City. In the circumstances envisaged in the example, Norman's reliance on his clairvoyant power can be seen as an 'effective', but not as a 'rational' course of

action for the sake of truth-acquisition and error-avoidance. However, Norman's present inability to provide a justification for his belief need not prevent us from attributing him the knowledge that the President is in New York City[3]. (I leave to the reader the easy task of devising a similar account for the hologram case, in which the problem is not the absence, but the invalidity of available epistemic justification).

Summing up, my claim is not that the distinction between knowledge attributions and knowledge claims frustrates *every* attempt to draw internalist conclusions from the hologram and clairvoyance examples. My claim is merely that such examples do not express unambiguously internalist intuitions about *knowledge*. For it is possible to devise an interpretation of those examples which preserves the basic intuition that (internal) epistemic justification is necessary to knowledge claims without subscribing to the more doubtful thesis that epistemic justification is a necessary, let alone a sufficient, condition for turning true beliefs into knowledge. Since a naturalistic conception of knowledge has the conceptual resources for dealing with such purported counterexamples, I conclude that our decision between internalism and pure (as Lehrer would say) or naturalistic (as I would rather call it) externalism in the analysis of knowledge will have to be grounded on stronger considerations than the hologram and clairvoyance cases alone.

5.3 'Animal' Knowledge

Foley (1987, 168) refers to the peculiar skill of chicken-sexers, who are claimed to be capable of determining the sex of very young chicks without being able to cite any distinguishing mark capable of explaining their ability, in order to defend the claim that in certain cases it may be plausible to think 'that a person knows *p* without [her] belief *p* being epistemically rational':

---

[3] For similar reasons, Putnam's (1982, 7) counterexample to what he calls the 'reliability theory of rationality' fails to provide a genuine counterexample to the externalist view of knowledge. Putnam's counterexample is as follows. Suppose that the Dalai Lama is infallible on matters of faith and morals. A person who believed everything the Dalai Lama had to say on matters of faith and morals would therefore acquire infallibly true beliefs on those subjects. But if that person's unique argument for believing in the infallibility of the Dalai Lama were 'the Dalai Lama says so', her method of belief-formation could hardly be described as 'rational'. Which is obviously true, but does not mean that that person's true beliefs on matters of faith and morals could not be described as genuine instances of theological and moral knowledge!

> Although [chicken-sexers] are able to distinguish male chicks from female chicks, they cannot describe to themselves or to others how they do this. Nor can they teach others to do it, for they do not have a conscious technique to teach.
> (Foley 1987, 168)

Apparently this is not just a story, nor are chicken-sexers merely pretending not to be aware of any marks distinguishing male chicks from female ones. But this is not really important. Even if there were in fact no chicken-sexers meeting Foley's description, the mere circumstance that there *might* be raises the issue, Would such a chicken-sexer *know* that the chick in her hand was female if she believed it to be so without having any inductive evidence of her own reliability as a chicken-sexer, and therefore without being able to provide any good argument in favour of her belief? My own intuition is that Foley's claim that 'such a chicken-sexer somehow, in a way that neither [s]he and perhaps no one else understands' would *know* that the chick in her hand was female is correct. But if so much is granted, then it is clear that someone can know that p even without being justified in believing that p. The case of chicken-sexers thus seems to me to support the claim that (internal) epistemic justification is not a necessary condition for knowledge. However, I reckon that intuitions about chicken-sexers may diverge. People like Bonjour will probably assimilate chicken-sexers to unsuspecting clairvoyants and add a new entry to their list of cases of mere possession of information. Therefore I will not rest all my case upon chicken-sexers.

A less controversial argument against the necessity of (internal) epistemic justification for knowledge is given by the existence of what we might call, in a very broad and non-derogatory sense, 'animal' knowledge, that is, the knowledge possessed by such epistemically unsophisticated subjects as animals and children. This point has been stressed, among others, by Alvin Goldman (1976), Alan Goldman (1988) and Dretske (1988). Its relevance for an externalist approach to knowledge is precisely that it can prompt a decision between those naturalistic analyses of 'S knows that p' that merely require that S have a true belief exhibiting the 'right' factual connection to the world and those quasi-traditionalist (but none the less externalist) analyses of 'S knows that p' that require that S be epistemically justified in believing that p as well as have a true belief exhibiting the 'right' factual connection to the world.

A young child can be described as *knowing* that her room is on the second floor, that she likes chocolate ice cream, that her brother's name is Peter, and so on.

Moreover, 'there is no intuitive reason to deny that a child knows such things as well as and in the same way that we all do' (Alan Goldman 1988, 38). Yet surely a young child cannot be described as being able to *justify* her beliefs that her room is on the second floor, that she likes chocolate ice cream, etc. For that matter, it is highly doubtful whether many philosophically unsophisticated adults would be able to justify claims about the location of their bedrooms or their ice cream tastes either. Shall we say that most people do not know the location of their bedrooms or their own ice cream tastes? It seems much more plausible to acknowledge that they do, and admit that (internal) epistemic justification fails to be a necessary condition for knowledge. This seems to me a strong reason for preferring a purely externalist (i.e., naturalistic) analysis of knowledge to those quasi-traditionalist accounts that require that S be epistemically justified in believing that *p* as well as have a true belief exhibiting the 'right' factual connection to the world.

Still, one might try to retain epistemic justification as a necessary condition for knowledge by distinguishing between being justified and showing justification: a young child's beliefs about her brother and chocolate ice cream might *be* justified without her being able to spell out the reasons of their justification. This line of reasoning, or something very close to it, is likely to be at work in many externalist accounts of epistemic justification, because it is clear that a justification which a subject's beliefs can possess without the subject's herself being able to spell it out is likely to be construed as an epistemically inaccessible (to the knowing subject[4]) justification.

However, I agree with Bonjour that clairvoyance cases like the one discussed in the previous section show that it is deeply counterintuitive to say that one can be epistemically justified in believing that *p and* have no epistemic access at all to one's own justification. The reason for this is, I suppose, that the concept of epistemic justification primarily refers to a person's *activity* of answering other people's challenges to her cognitive claims by *justifying* her making those claims. (One might draw a parallel with ethical justification and point out that it is actions, not states — claims, not beliefs — that primarily require justification). S's capacity of justifying the claim that *p* against other people's challenges can then be said to justify S's *believing* that *p*. Finally, S's

[4] Hard-line externalists will maintain that S's epistemic justification for (truly) believing that *p* need not be accessible to *anyone* for S to know that *p*. But it might be plausible to claim that even if S's epistemic justification for (truly) believing that *p* need not be epistemically accessible to S for S to know that *p*, still there must be some subject other than S to which it *is* (in some sense) epistemically accessible.

belief that *p* can itself be described as justified, but it seems to me that since the concept of (epistemic) justification applies to the status acquired by S's belief that *p* only in virtue of S's being justified in believing that *p*, a belief can be said to be justified only in a derivative way. If this is true, it is easy to perceive the peculiarity of a justified belief which has not actually been justified and the possessor of which is wholly incapable of justifying. (An external observer may describe the belief in question as justified because, say, it has been produced by a cognitive method she knows to be reliable; but it seems clear to me that the mere fact that the cognitive method *is* reliable cannot justify S's *believing* that *p*, let alone S's *claiming* that *p*. Therefore the fact that an external observer may classify S's belief that *p* as justified on the ground that it has been produced by a reliable cognitive method does not seem to me to undermine the intuition that one cannot be epistemically justified in believing that *p and* have no epistemic access at all to one's own epistemic justification. What the external observer means by saying that S's belief that *p* is justified is probably that it has been produced in such a way that it is possible to justify the claim that *p* by appealing to the process that has brought about S's believing that *p*, or, in other terms, that S's believing that *p* carries the information that *p*).

Of course ordinary language is not sacrosanct, and one could simply stipulate to call 'epistemically justified' not only a cognitive claim or belief for which an epistemic justification has been provided, but also, say, a belief which has been produced by a reliable belief-forming process or which is a reliable indicator of the state of the world. But I think that insisting on calling such an externalist relation 'epistemic justification' simply because it is proposed as a (third) condition in the analysis of knowledge would be misleading, because there is virtually nothing in common between 'external' and 'internal' epistemic justification, as it will become clear in the remainder of this chapter. However, since our present concern is to reject the necessity of *internal* epistemic justification for knowledge in order to defend a purely externalist view of knowing, the possibility of an externalist interpretation of epistemic justification is irrelevant to our problem, since in any case it would fail to affect the viability of an externalist account of knowledge.

That (internal) epistemic justification is unnecessary for knowledge is also suggested by more literal cases of 'animal' knowledge, such as an ant *knowing* that forage is available at the end of the tracks of the foragers who have come back heavy

laden, or a hen *knowing* that the approaching fox represents a danger. As Dretske (1981, 209 ff.) has pointed out, one may be reluctant to ascribe beliefs to animals, especially because they have no language of their own, which makes it rather difficult to capture the precise content of their beliefs in *our* language. Does John's cat believe that *s* is liver, or does it merely believe that it is food? Does Ann's dog believe that *s* is the postman, or does it merely believe that he (it?) is an intruder? Nevertheless, while it is certainly true that one must be very careful in describing the content of animal beliefs, this does not prove that there is anything wrong in our habit of ascribing knowledge to animals:

> A fundamental facet of animal life, both human and infra-human, is telling things apart, distinguishing predator from prey, for example, or a protective habitat from a threatening one. The concept of knowledge has its roots in this kind of cognitive activity.
>
> (Alvin Goldman 1976, 791)

Alvin Goldman writes that we may tempted to use the concept of 'knowledge' even to describe cases of 'mechanical' information-processing, such as an electric-eye door *knowing* that someone (something) is coming. He is careful to qualify his statement by saying that in such cases the concept of knowledge is used somewhat metaphorically or analogically, but he claims that a correct definition of knowledge 'should be able to explain extended and figurative uses as well as literal uses, for it should explain how speakers arrive at the extended uses from the central ones' (Alvin Goldman 1976, 791). And he emphasizes that a (purely) externalist account of knowledge satisfies this requirement, while an internalist one doesn't.

I do not want to rest too much weight on the fact that purely externalist (naturalistic) accounts of knowing can easily handle mechanical cases of information-processing, because it seems clear that although relatively simple systems, such as a thermometer, can be described as *representing* external states of affairs, they can hardly be described as having *beliefs* (and hence knowledge) about the states of affairs they represent. For a belief is not just a representational structure, but a part of a larger representational manifold including desires, emotions, intentions and attitudes as well as beliefs.

Indeed, the distinction between beliefs and simpler representational structures may throw some light upon the difference between knowledge and (mere) possession of

information. While S's knowing that *p* involves S's *believing* that *p*, S's possessing the information that *p* merely involves S's being able to *represent* that *p*; knowing that *p* involves possessing the information that *p*, but not *vice versa*. From the standpoint of the information-theoretic approach to knowledge discussed in the last chapter, I would claim that what turns (mere) possession of information into knowledge is the functional role played by S's representation that *p* in the larger representational manifold that turns S's representation that *p* into S's *belief* that *p*. A virtue of this way of drawing the distinction between knowledge and mere information-possession is that it agrees with the fact that S can possess the information that *p* without believing that *p* — that this can be so is evident from the fact that inanimate representational systems are capable of processing information, as well as from the fact that human beings can fail to remember having acquired the information that *p* (which seems to me the right conclusion to draw from the third story told by Radford 1966), or fail to recognize that *p* is entailed by other propositions they know to be true.

This way of drawing the line between knowledge and mere possession of information will of course fail to satisfy those philosophers who claim that naturalistic accounts of knowledge are hopelessly misguided because any analysis of 'S knows that *p*' which fails to require that S be internally justified in believing that *p* cannot be anything more than an analysis of 'S possesses the information that *p*'. But again, all the evidence they have for claiming that naturalistic analyses of knowledge merely succeed in capturing the notion of information-possession is given by that sort of examples I have argued cannot be taken to express unambiguously internalist intuitions about *knowledge*. The picture arising from the conjunction of the last paragraph's distinction between knowledge and mere information-possession with the view that (internal) epistemic justification, although unnecessary for knowledge, is necessary for knowledge claims seems to me to provide a plausible account of the intuitions about knowledge and epistemic justification elicited *both* by the hologram and clairvoyance examples *and* by cases of 'animal' knowledge. On the other hand, although the view that (internal) epistemic justification is necessary for knowledge appears to do justice to the intuitions elicited by the hologram and clairvoyance examples, it fails to account for those instances of knowledge in which the knowing subject is obviously incapable of providing a justification for her beliefs.

The existence of 'animal' knowledge represents a problem for all those (internalist or externalist) overintellectualized accounts of knowing that view internal epistemic justification as a necessary condition for knowledge. If one wants to stick to the idea that internalist epistemic justification is a necessary condition for knowledge, one will have to dismiss as irrelevant all those knowledge-attributions which have animals or young children (and epistemically unsophisticated adults) as their subjects, and will end up inventing new epistemic categories (e.g., 'proto-knowledge') to account for the persistency of those attributions in our ordinary descriptions of the world (see Alan Goldman 1988, 38). This may not be a knock-down argument against the recommendation that internal epistemic justification be regarded as a necessary condition for knowledge, but it brings out a difficulty that a naturalistic account of knowing surely need not confront.

5.4 Justification and Normative Epistemology

Although internal epistemic justification can very well be regarded as unnecessary to knowledge, cases like that of the unsuspecting clairvoyant suggest that it is necessary to back knowledge *claims*. We have already noted that this fact is often used to substantiate the complaint that (purely) externalist accounts of knowing, failing to count internal epistemic justification as a requirement for knowledge, have nothing to say about how we should behave in order to improve our cognitive situation. In other words, (purely) externalist accounts of knowing are alleged to be worthless for the purpose of developing a *normative* epistemology.

This criticism is, I think, mistaken. A reliabilist account of knowing, for example, may very well recommend that we use reliable belief-forming processes in our cognitive efforts. For the fact that a belief-forming process can be reliable without being known or justifiedly believed to be so does not entail that *no* reliable belief-forming process can be known or justifiedly believed to be reliable. There appears to be no reason why a purely externalist account of knowledge should have nothing to say about how we should behave in order to improve our cognitive situation. It is not externalist knowledge that stands in the way of a normative epistemology, but the claim that our beliefs can never enjoy anything but externalist *justification*, and it is only if the proposal

of an externalist account of knowing is *ipso facto* interpreted as the proposal of an externalist account of epistemic justification that the former can appear to hamper the development of a normative epistemology.

Here is a very simple argument to the effect that an *internalist* concept of epistemic justification is necessary to the development of a normative epistemology (see, e.g., Alvin Goldman 1980, 28, and Nagel 1986, 69).

A normative epistemology will typically advise cognizers to believe (accept) exactly those beliefs which come out as (comparatively) best justified on the basis of its own criteria of justification. But if one is to follow this suggestion, one will need to have epistemic access to the *justifiedness* of one's own beliefs. A suggestion like S: 'Retain justified beliefs and reject [suspend judgement on] unjustified beliefs', will be completely idle if one is not in a position to tell which of one's own beliefs should be counted as epistemically justified and which should not. Accordingly, the notion of justification can be assigned a 'normative' role in our epistemology only if justifiedness is taken to be epistemically accessible, that is, if an internalist notion of epistemic justification is adopted.

I have already said that I do not think one should endorse an externalist notion of epistemic justification merely because one needs a name for an externalist 'third condition' for knowledge. The argument that has just been sketched is, however, a bit too quick, at least if an externalist notion of epistemic justification is not taken as excluding a priori any possibility of gaining epistemic access to the justifiedness of some at least of one's own beliefs. Again, one could point out that the fact that a belief can be externally justified without being known or justifiedly believed to be so does not entail that *no* externally justified belief can be known or justifiedly believed to be justified. Apparently, nothing prevents one from developing a normative epistemology on the basis of an externalist notion of epistemic justification by replacing S with S*: 'Retain those beliefs that you know [justifiedly believe] to be justified and reject [suspend judgement on] the others'.

However, it is clear that the cognitive policy recommended by S* would be completely useless if no beliefs could in fact be known [justifiedly believed] to be epistemically justified. For if no beliefs could be known [justifiedly believed] to be epistemically justified, S* would be vacuous (if one were to try to substitute mere belief for knowledge [epistemic justification] in S*, S* would certainly cease to be vacuous, but

one would be left with no reason at all to believe that following S* should lead to an improvement in one's epistemic situation). The point is that if someone wanted to follow recommendations like S* while having no epistemic access at all to the justifiedness of her own beliefs, she would not know where to start from. This is the reason why internal epistemic justification appears indeed to be essential for the development of a normative epistemology. It is not because one could not formulate an equivalent epistemology in externalist jargon, but because if externalist justification was the only sort of justification enjoyed by our beliefs, it would be impossible to follow the recommendations of any normative epistemology whatsoever. Therefore I think one ought to conclude that there must *be* internally justified beliefs for a (successful) normative epistemology to get under way at all.

This conclusion is not, however, as significant as it may appear. For the assumption that the formulation of a 'Grand Normative Theory' that would recommend optimal strategies for the cognitive efforts of human beings is a meaningful and viable enterprise is itself open to question. The possibility of devising such a theory is being denied by an increasing number of naturalistically-minded philosophers. If the task of normative epistemology is to provide content-independent decision principles for all or most epistemic problems human cognizers can encounter, I believe (and I shall argue in the next chapter) that these philosophers are basically right. But even if they were simply mistaken and the project of devising such a Grand Normative Theory were not pointless, there seems to be very little evidence that it has as yet been significantly successful or that it is likely to meet with any significant success in the foreseeable future. The starting-point of the argument from normative epistemology would then appear to be a possibility which is as yet only very partially and imperfectly realized in the cognitive efforts of human beings.

A way to circumvent this sort of problem is to run the argument not from the existence or possibility of a Grand Normative Theory recommending optimal strategies for the cognitive efforts of human beings, but from the plain fact that human beings do, after all, engage in the activities of making, evaluating and criticizing knowledge claims, and in so doing appeal to several kinds of cognitive principles. This apparently undeniable fact leads me to the topic of the next section, which is the necessity of internal epistemic justification for an account of successful argumentative interaction between human beings.

5.5 Justification and Successful Argumentative Interaction

The argument of this section purports to show that if there is anything like successful argumentative interaction between human beings, then there must be internally justified (as well as internally unjustified) beliefs. This is so because it is only the *internal* justification of beliefs that has a direct bearing on the success of our argumentative attempts to persuade other people to believe this and that. (By 'successful argumentative interaction' I mean that kind of linguistic interaction which is not merely rhetoric and yet is capable of leading to consensus. It is beyond the scope of this work to prove that human arguments can indeed be divided into valid and invalid, and not merely into persuasive and unpersuasive. But surely the belief that they can is not exclusive of upholders of correspondence truth).

I will discuss an example of (unsuccessful) argumentative interaction which will show, I hope, that it is possible to make sense of the activities of making, evaluating and criticizing knowledge claims without seeing them as applications of a Grand Normative Theory recommending optimal strategies for our cognitive efforts. If my argument is sound, the concept of epistemic justification which is required to account for the possibility of such activities need not be construed as having any internal or a priori connection with truth: the rejection of the traditional project of devising a Grand Normative Theory makes it unnecessary to build the requirement that a justified belief be likely to be true into the very concept of epistemic justification.

This of course is not to deny that truth plays a crucial role in the definition of the goals of our cognitive efforts and that we shall typically do our best to ensure that those beliefs which come out as epistemically justified on the basis of our criteria of justification are more likely to be true than false. But our best efforts may not be good enough and our criteria of justification may fail to be adequately truth-conducive. Therefore I think that we are well advised not to conflate the issue of the epistemic status of a given belief with respect to a certain set of criteria of justification with the issue of the truth-conduciveness of those criteria. Incidentally, Chisholm (1989, 76) takes the thesis that there is no logical connection between a belief's being epistemically justified and its being likely to be true as an hallmark of internalism[5]. This thesis is,

[5] Chisholm writes in fact that internalism is committed to the claim that 'there is no *logical* connection between epistemic justification and truth. A belief may be internally justified and yet be *false*'.

however, perfectly consistent with those brands of knowledge-externalism that do not take external epistemic justification to be a necessary condition for knowledge. Indeed, it seems to me that the thesis that there is no logical connection between a belief's being epistemically justified and its being likely to be true sits much less comfortably with an internalist theory of knowledge than with an externalist one. Bonjour (1980, 54) seems to me to be right when he claims that according to traditional internalist epistemologies a cognitive act is epistemically justified only if it involves acceptance of 'beliefs that there is adequate reason to think are true'. Because of the rejection of this thesis, Chisholm's theory of knowledge turns out to be in fact nothing more than a theory of epistemic justification, as no effort is made to explain why justification should represent an epistemic virtue with respect to the cognitive goal of the search for truth. The absence of any reference to truth in Chisholm's treatment of epistemic justification has the unpalatable consequence that his epistemology, albeit undoubtedly normative, cannot be seen as providing any reliable advice for the pursuit of truth, as opposed to the pursuit of epistemic justification.

I shall describe every view of epistemic justification that postulates, à la Bonjour, an internal or a priori connection between epistemic justification and truth as providing a 'thick' conception of epistemic justification. And I shall describe every view of epistemic justification that fails to do so as providing a 'thin' conception of epistemic justification. I introduce this distinction because it is my purpose to emphasize that the only sort of internal epistemic justification the activities of making, evaluating, and criticizing knowledge claims can be seen as providing unequivocal evidence for is *thin* epistemic justification.

Even though Norman's inability to provide a justification for his true belief that the President is in New York City should not be taken as evidence that he does not know the whereabouts of the President, it surely prevents him from persuading anyone else of

---

Of course saying that an internally justified belief may be false does not commit one to deny that internally justified beliefs are likely to be true. But for all my efforts, I have not been able to locate in Chisholm's *Theory of Knowledge* any explicit admission of the existence of a probabilistic connection between justification and truth. To be sure, Chisholm appears to endorse the claim that 'autopsychological' statements are certain, and thus, one may suppose, certainly *true* (see Chisholm 1989, 22-25). Yet his treatment of certainty is, as his treatment of epistemic justification, officially quite separate from any question of truth. Certainty and epistemic justification are treated as epistemic notions, and no mention is made of their bearing upon truth.

That Chisholm's epistemology cannot be seen as recommending maximally truth-conducive strategies for the cognitive efforts of human beings is also argued, from a different perspective, by Goldman (1980, 41-42).

the truth of what he believes. But if we want to get a better understanding of his predicament, I think that a different story will prove more instructive.

Mary and John are, unsurprisingly, two philosophy students. Both of them truly believe — indeed, we can suppose that they know — that q, that p entails q, that $\neg q$ entails $\neg p$, and that p provides the best available explanation of q. Having read van Fraassen (1980) and Lipton (1991), neither of them thinks that explanations must be true. Yet Mary believes that 'inference to the best explanation' (henceforth: IBE) preserves truth, while John does not. So Mary believes that p, while John remains agnostic. We can further suppose that, unknown to Mary (and John), p is not only the best available explanation of q, but (in some sense) the only possible explanation of q. So it may be the case that Mary came to believe that p by means of a reliable belief-forming process (because in this particular case the best available explanation of q happens to be in fact the only possible explanation of q, so that the truth-value of q may turn out to be a reliable indicator of the truth-value of p). Under such circumstances, a partisan of external justification would no doubt be inclined to say that Mary's belief that p, being the result of a reliable belief-forming process, is *externally* justified.

But will the fact that Mary's belief that p can be described as externally justified provide her with an argument to convince John that p? The answer is, of course, that it will not. But if we want to give an *account* of why Mary cannot persuade John to share her (externally justified) belief that p, we will have to use an *internalist* concept of epistemic justification in our description of their argumentative situation: Mary cannot persuade John to believe that p because they do not share the same criteria of *internal* justification — Mary believes that IBE provides justification, while John believes that it does not. Accordingly, Mary believes that her belief that p is internally justified, while John believes that it is not. If John also believed in the reliability of IBE, Mary could convince him that p by appealing to an argument of that form. But since he does not believe in the reliability of IBE, there is no way (short of persuading him of its epistemic virtues) in which Mary can present him with an argument to the effect that p.

This example is admittedly scarcely representative of real argumentative situations because IBE is a very abstract and content-independent principle, whereas the cognitive principles appealed to in real argumentative situations are usually not so. This said, the example does show, I think, that the extent to which two cognitive subjects can engage in a successful argumentative interaction depends (among other things) on what their

conceptions[6] of internal justification for beliefs on the matter at issue are. By this I mean that it depends (and this again is an idealization if compared to real argumentative situations) on the particular *sets of criteria* of epistemic justification the two subjects happen to endorse for beliefs on the matter at issue (a conception of epistemic justification can be described *explicitly* by providing a list of the members of its defining set, or *implicitly* by providing a list of cognitive principles which prescribe exactly those beliefs that satisfy the members of its defining set). The extent to which a cognitive subject can engage in a successful argumentative interaction with another depends on the relationship between their conceptions of what can *recognizably* confer justification upon what with respect to the matter at issue. Everything else being equal, the better the match between their conceptions of internal justification with respect to the matter at issue, the higher the chances for a successful argumentative interaction to take place. Two cognitive subjects who lacked any conception of internal justification would be incapable of engaging in any form of argumentative interaction, because they would lack any *argumentative* (as opposed to physical and rhetorical) means of persuasion. In the case of our unsuspecting clairvoyant, there is no reason to believe that Norman lacks any conception of internal justification, but whatever conception of internal justification he has, it is not satisfied by his belief that the President is in New York City. Therefore, he cannot appeal to any shared standards of justification to persuade other people of the truth of his belief.

One might think, though, that our IBE example merely shows that Mary and John have different conceptions of internal epistemic justification, not that there *are* internally justified (as well as internally unjustified) beliefs. After all, what the example shows is only that *p* is *regarded* by Mary as internally justified and by John as internally unjustified. Why should there be any fact of the matter as to whether *p is* internally justified or not? Could not Mary and John merely be deluding themselves about the existence of internally justified and unjustified beliefs? This objection would of course have some bite if we were working with a *thick* conception of epistemic justification. For the fact that *p* is regarded by Mary as internally justified and by John as internally unjustified obviously does not show that there is a set of optimal cognitive principles

[6] I write 'conceptions' rather than 'concepts' because two persons can share the same concept of epistemic justification while having different opinions ('conceptions') about what can confer justification upon what (the plausibility of this distinction is argued, e.g., in Putnam 1981, ch. 5).

according to which p is either internally justified or internally unjustified. But the result of our argument is to be construed in terms of *thin* epistemic justification. On this construal, being (or failing to be) internally epistemically justified boils down to recognizably satisfying (or failing to satisfy) a given set of criteria of justification. What the argument shows is then that the existence of successful argumentative interaction between human beings requires the existence of beliefs which recognizably satisfy *particular sets of criteria* of justification.

This way of looking at epistemic justification will perhaps seem less eccentric if one pays attention to the fact that the history of Western thought, including the history of post-Galilean science, as well as several ethnographical studies, suggest that there have been in fact different concepts of epistemic justification adopted by different human groups, diachronically as well as synchronically. It is the existence of shared concepts of epistemic justification, not their truth-conduciveness, that accounts for the possibility of successful argumentative interaction between human beings. Therefore there seems to be a *prima facie* case for treating the concept of epistemic justification as a relative concept, in the sense that at any one time a given belief or knowledge claim may be both J_1-justified and J_2-unjustified, where J_1 and J_2 are non-equivalent (and possibly unequally truth-conducive) sets of criteria of justification. My proposal is then to treat the question of the J_i-justification (or lack of justification) of S's belief that p on the one hand, and the question of the *vindication* of the set J_i of criteria of justification on the other, as independent issues (this approach is very close to the 'epistemological relativism' sketched in Field 1982, 562-567)[7].

---

[7] The recognition of the relativity of the concept of epistemic justification might suggest a strategy for rejecting the argument from 'animal knowledge' expounded in section 5.3 above. By granting the relativity of epistemic justification, one could point out that the incapacity of unsophisticated adults to justify their beliefs about the whereabouts of their bedrooms to the satisfaction of epistemologists is no evidence that they do not know where their bedrooms are for the simple reason that the sort of (internal) justification their beliefs must enjoy for being counted as knowledge is to be assessed with reference to different and less stringent criteria than are currently adopted in epistemological circles.

It seems to me that this rejoinder will be of little help to those who see internal justification as a necessary condition for knowledge. Apart from the obvious fact that developing this rejoinder commits one to give up seeing epistemic justification as that property that turns true beliefs into knowledge, the trouble is that for most (true) beliefs we would be disposed to regard as knowledge, there is no evidence at all that their possessors would be capable of providing even a rudimentary justification of their content. It is not as if they were incapable of justifying their beliefs before a philosophical court. What they are incapable of doing is justifying them at all — indeed, it is doubtful if they would understand the point of a request of epistemic justification.

The grain of truth that can be found in epistemological internalism is then that the existence of *thinly* internally justified beliefs, though unnecessary for knowledge, is necessary for our argumentative practices, which include the activities of making, evaluating and criticizing knowledge *claims*. So much can be granted without abandoning the naturalistic approach to knowledge described in chapter 4, which appears to be fully consistent with the intuitions expressed by the hologram and clairvoyance 'internalist' stories. Indeed, if the argument presented in section 3.4 is sound, it is an *internalist* approach to knowledge that appears to sits rather uncomfortably with an internalist account of epistemic justification. And this means that if the main reason that leads fallibilist internalist epistemologists to subscribe to an internalist view of knowledge is their believing that an internalist view of knowledge is entailed by an internalist view of epistemic justification, the foundations of knowledge-internalism begin to appear very shaky indeed. But to go back to the distinction between thin and thick justification, my point is that the support granted by the hologram and clairvoyance 'internalist' stories to *thin* justification-internalism does not extend to *thick* justification-internalism because those stories provide no evidence whatsoever for the claim that our argumentative practices *are* truth-conducive, or for the claim that the activities of making, evaluating and criticizing knowledge claims only make sense as applications of a Grand Normative Theory recommending optimal strategies for our cognitive efforts. Therefore arguing that we *need* (thick) internal justification on the ground that we *need* a Grand Normative Theory seems a scarcely viable enterprise.

The relevance of the (true) claim that the existence of internally justified beliefs is necessary for the project of normative epistemology ought then not to be overestimated. If the claim is merely that the existence of internally justified beliefs is necessary for making *prescriptions*, then of course it will only support the existence of *thinly* justified beliefs. If, on the contrary, the claim is meant to be that the existence of internally justified beliefs is necessary for making *truth-conducive* prescriptions, the hologram and clairvoyance 'internalist' stories will provide no evidence that we know how to make truth-conducive prescriptions in the first place.

Genuinely argumentative interaction between human beings is perhaps less common than one might wish to think, but its existence is a widely recognized fact, which is denied only by those thinkers who either emphasize the pervasiveness of power-relations in all aspects of human life, or believe that arguments do not divide into good

and bad, but only into persuasive and unpersuasive. On the other hand, the project of normative epistemology raises serious questions. While the existence of successful argumentative interaction among human beings only requires the achievement of consensus, the development of a normative epistemology requires the identification of truth-conducive methodological principles, which is a much trickier matter.

Summing up, I take it that the arguments of this section have shown (i) that since one cannot affect argumentatively other people's beliefs without inducing them to *see* the reasons why a change in their system of beliefs is being required, there must be internally justified (as well as internally unjustified) beliefs; (ii) that this holds true whether the reasons one appeals to are 'good' reasons or 'bad' reasons; and (iii) that it is perfectly possible to make sense of the fact that people engage in the activities of making, evaluating and criticizing knowledge *claims* without endorsing an internalist account of *knowledge*.

5.6 The Pragmatic Character of Epistemic Justification

Earlier in this chapter I noted that the concept of epistemic justification primarily refers to a person's activity of answering other people's challenges to her cognitive claims, and only derivatively applies to the status acquired by a belief in virtue of its being held *justifiably*. In this section I propose to explore in greater detail the consequences of this fact. But before doing this, I want to dispel fears that an internalist view of epistemic justification might be exposed to the very kind of epistemic regress that led to the rejection of knowledge-internalism in section 3.4 above.

As a matter of fact, a similar regress does affect some forms of justification-internalism, for example a justification-internalism construed as requiring that S *justifiedly believe* that the conditions which must be met if a belief is to be counted as justified are indeed satisfied. But having adopted an externalist view of knowing which does not make of epistemic justification a necessary condition for knowledge, I am now free to interpret justification-internalism as requiring that S *know* that the conditions which must be met if a belief is to be counted as justified are indeed satisfied. This does not seem to be an exceedingly strong requirement for someone committed to understanding internal epistemic justification in the terms of shared sets of

methodological rules rather than as that (relational) property that turns true beliefs into knowledge. And since our concept of knowledge is externalist, requiring that S *know* that the conditions for epistemic justification are satisfied will not bring about any regress. A carefully formulated version of justification-internalism can thus be trusted not to be threatened by dangers of epistemic regress.

This said, we can now turn to the main subject of this section, which is the pragmatic character of epistemic justification. If internal epistemic justification is regarded as the way in which our beliefs are epistemized, then of course it must be taken to be at least as objective and context-independent as knowledge is taken to be. But if internal epistemic justification is regarded as an activity meant to affect other people's (and one's own) beliefs, it will come as no surprise to discover that it is a strongly context-dependent activity. This is because it inherits the context-dependence of the knowledge *claims* it is meant to support. If it is true that the primary objects of justification are concrete knowledge claims and that beliefs may be said to be justified only in a somewhat derivative way, what may count as a justification of the belief that *p* in a certain context may fail to do so in different circumstances.

When I produce to my Italian friend Marco, who is happily ignorant of British politics, a recent copy of *The Times* in order to justify my claim that Mr. Clarke, rather than Mr. Major, is the present Chancellor of the Exchequer, I am not trying to establish the truth of the proposition, 'Mr. Clarke is the Chancellor of the Exchequer on 14 February, 1994'. What I am really trying to do is to appeal to Marco's belief that *The Times* is a reliable source on British affairs to argumentatively persuade him of the truth of my claim. Alternatively, I could attempt to persuade him by producing a copy of *The Independent*, but since my friend Marco has never heard of any reliable British newspaper but *The Times*, and furthermore suspects that *The Independent* could be a satirical paper (he is discouraged from taking it seriously by the untrustworthiness of *L'Indipendente* Italian newspaper), I would fail to justify my claim that Mr. Clarke is the present Chancellor of the Exchequer. *The Independent* might succeed in convincing Ilaria, who would read it while she was in Britain two years ago, but surely will not convince Marco.

The point made by this example is not that, if Marco had never heard about the reliability of *The Times*, I would have been unjustified in claiming that Mr. Clarke is the present Chancellor of the Exchequer. Presumably, if Marco had never heard about the

reliability of *The Times*, I could have appealed to some further background knowledge of his to make him see that my claim was indeed justified. The point is rather that the activity of epistemic justification need not be carried on until all conceivable doubts about the truth of the relevant proposition have been dispelled. On the contrary, it comes to an end as soon as it has achieved its pragmatic goal of defeating the epistemic challenge raised by the objector, that is, as soon as the objector has been led to accept the truth of the claim which is being justified. In this sense, I contend that the epistemic justification we actually provide for our knowledge claims is in a strong sense context-relative. (This feature of epistemic justification may help to understand why epistemic justification in real, interpersonal contexts is possible even if epistemological foundationalism is untenable).

The beliefs of the person to whom the justification is addressed appear to determine what can count as a justification of a knowledge claim in a given context. If the aim of epistemic justification is not to epistemize but to affect beliefs, I suggest that the kind of epistemic justification one commits oneself to provide to one's audience when uttering the claim that p is precisely this kind of context-relative epistemic justification. Since none of us is engaged in the Cartesian 'project of pure enquiry', asserting that p does not commit one to establishing the truth that p from scratch, but merely to providing arguments that may persuade one's intended audience to believe that p.

An audience (which I shall assume to be homogeneous) is characterized, roughly, by a set of beliefs about the world and the nature of epistemic justification as well as by the range of evidence to which it has (direct or indirect) access. These features affect what can count as a justification of a knowledge claim for a given audience: e.g., Marco's belief about the reliability of *The Times* enables me to justify the claim that Mr. Clarke is the present Chancellor of the Exchequer by producing a copy of *The Times*, rather than *The Independent*, before him.

I propose to say that S is justified in making the claim that p before audience A iff S would be able, if her claim were to be challenged, to provide A with an epistemic justification of p. Here the primitive notion is that of an *act* of epistemic justification: S's knowledge claim that p is said to be justified before audience A if S has actually provided an adequate justification for it, and S is said to be justified in making the claim that p before audience A if she would be able, on request, to perform such an act.

According to this notion of justification, a knowledge claim, and derivatively a belief, can indeed be justified, but only *relative to a certain context* (i.e., relative to the beliefs and evidence available to a certain audience). This context-relative justification is undoubtedly *epistemic* justification, because it is intended to answer challenges about the *truth* of the asserted proposition: it is not intended, for example, to answer challenges about the rightness of performing a given action. This is an important point: even though the possibility of successful argumentative interaction does not require that epistemically justified beliefs be likely to be true, it does require that criteria of epistemic justification be selected in view of their cognitive, and not, for example, rhetorical, effectiveness (more about this in the next chapter). But I say that this context-relative justification has, unlike knowledge, a *pragmatic* character because it is not the sort of epistemic justification a Cartesian 'pure enquirer' could be satisfied with. S's knowledge claim that *p* can be justified because it is not itself an abstract proposition, but a linguistic *act* addressed to a particular audience. And S's belief that *p* can inherit such a justification and be described as justified *before that particular audience*, which means that in general the propositional content of S's belief cannot be said to be 'absolutely' justified, independently of the intended audience of S's claim that *p*.

It is true, however, that some of our knowledge claims tend to transcend the actual situation in which they are made, exhibiting an aspiration to a more general validity. This is typical of scientific (and, maybe, philosophical) claims. These claims are usually meant to be justified not merely before this or that audience, but before any audience which fulfils certain general requirements of rationality. Beware: this does not mean that in more ordinary cases the proposition which is claimed to be true is not claimed to be true *for everyone, everywhere and in every time*. What is lacking in more ordinary cases is rather the commitment to *justify* the claim before every rational subject everywhere and in every time. Even in the case of science, however, it is clear that this commitment involves a considerable degree of idealization, for no real audience is likely to be a perfectly rational audience. Still, the methodology of scientific knowledge is arguably our best shot at the de-contextualization of epistemic justification. Scientific methodology sets severe constraints upon the accessibility of scientific evidence, so that the context of validity of scientific justification may coincide with the context in which the standards of scientific methodology are actually accepted. Scientific evidence must be public and open to inter-subjective testing. Revelations, feelings of conviction,

intuitions, and other private psychological states do not qualify as scientific evidence, because only a few privileged individuals can have access to them. The only constraint on the possibility of carrying out scientific justification is thus connected with the acceptance of scientific methodology, because the universal accessibility of the relevant evidence is built into the very definition of 'scientific' evidence[8]. And we shall see in the next chapter that scientific methodology itself is not understood as a 'take it or leave it' matter, but as liable to rational evaluation.

It is worth emphasizing that saying that there is a pragmatic dimension to epistemic justification is different to saying that there are no criteria of epistemic justification. Once the context (audience) is fixed, it is an objective matter whether S's claim that *p* is epistemically justified, because it is an objective matter what can count as an epistemic justification of S's claim that *p* before the relevant audience. Moreover, the recognition of the context-relativity of epistemic justification will not support, in the present framework, relativistic arguments against the possibility that human beings may have developed a context-independent knowledge of the world. Since epistemic justification is no longer seen as a necessary condition for knowledge, there is no reason why the latter should be seen as inheriting the context-relativity of the former. This is perfectly clear in the cases of animal and unreflective knowledge, where the question of epistemic justification does not even arise. But it is also clear for more sophisticated instances of knowledge, where it is surely plausible to think that a context-relative *activity* may lead to the development of context-independent *knowledge*, if this is conceived in the naturalistic fashion described in the last chapter.

5.7 A Non-Psychologistic View of Knowledge

In this chapter I have tried to argue that a proper blend of knowledge-externalism and justification-internalism can provide a perfectly plausible account of the 'internalist'

---

[8] Karl Popper laid much emphasis on this point (see Popper 1959, section 8), ending up endorsing a conventionalist view of the acceptance of the basic statements of science (see Popper 1959, 106). If the goal of epistemic justification is to *epistemize* beliefs, Popper's conventionalism on basic statements will have disruptive consequences for his falsificationist methodology (how can a decision refute a theory?). But if the goal of epistemic justification is to *affect* beliefs, it may be perfectly rational to rest the outcome of a discussion on a previous agreement about a few shared beliefs. Furthermore, if those shared beliefs happen to be true, the outcome of the discussion may be not merely consensus, but truth.

intuitions embedded in such purported counterexamples to a purely externalist view of knowing as the hologram and clairvoyance cases. This account involves acceptance of a pragmatic theory of (internal) epistemic justification, according to which beliefs can be said to be justified only in virtue of the logically prior activity of justifying concrete knowledge claims in real, interpersonal circumstances. I do not pretend to have shown that this is the only plausible view of epistemic justification, and I am aware that many readers will find an 'absolutist' view more congenial. These readers I ask to have some patience. In the following chapter I shall address the particular version of the issue of the vindication of methodological rules which one has to confront if (as I tried to argue in chapter 3) knowledge-internalism is untenable. I hope that discussion will contribute to clarifying what can realistically be required from a theory of epistemic justification, showing that a pragmatic theory of epistemic justification is not, after all, completely unsatisfactory. In the meantime, I want to add a few words on the anti-psychologistic role played by the correspondence theory of truth in the epistemological picture I am trying to sketch.

Correspondence truth provides us with a sense in which a sentence can be said to be objectively true (or false) independently of the epistemic warrant one may have for accepting (or rejecting) it. If what makes a sentence either true or false is its structure, the referential relations between its parts and reality, and the knowledge-independent nature of that reality (see section 1.4 above), then the issue of the truth-value of a sentence will be totally distinct from any psychological issue concerning the degree of confidence to which a given subject is willing to believe it, as well as from the epistemic issue of the justification enjoyed by a given knowledge claim having that sentence as its content. I may believe that a given sentence is true, I may be justified in claiming that it is true, and yet, according to the correspondence theory of truth, the truth-value of that sentence will not depend on my belief or on the grounds of my belief. That sentence's being true will never be identical with, or a function of, its being *justifiedly believed* to be true.

If truth is no longer seen as a function of some epistemic concept, there can be knowledge without epistemic justification. This fact makes it possible to understand knowledge as a real, factual relationship between a subject and her environment. The obtaining of such a relationship becomes a completely objective matter, which has nothing to do with the knower's capability of providing a ground for her own beliefs.

The lack of any conclusive, or even probable, justification of the knower's beliefs is wholly compatible with her being in possession of some real knowledge[9]. Such a purely externalist view may be seen as the most consequential form of an anti-psychologistic conception of knowledge.

In its classical formulation in the writings of Frege and Husserl, anti-psychologism represents a reaction to the empiricist identification of the laws of logic with the laws of human thought. Both philosophers took great pains to persuade their contemporaries that the truths of logic must not be understood as empirical generalizations about what human beings can (or cannot) believe, and as a whole they succeeded in their effort. Both Frege and Husserl were prompted to oppose psychologism in the philosophy of logic by their objectivism about truth: they could not accept the idea that (logical) truth could be in any way dependent upon the judging subject. Making logical truth dependent upon the judging subject would have meant relativizing it to the contingent (and possibly variable) constitution of the human species. Being reluctant to accept such a relativization, Frege and Husserl strenuously argued that logical truths, far from being empirical generalizations about mental processes, must be regarded, in Frege's (1967, 13) phrase, as 'boundary stones set in an eternal foundation, which our thought can overflow, but never displace'.

Even if I think that there still are good reasons to subscribe to an anti-psychologistic view of logic, today we are unlikely to be so dogmatic about the status of logical truths. However, such an externalist view of knowledge as I am trying to defend seems to me to provide a natural complement for an objectivist view of truth, and represents the most thoroughgoing outcome of an anti-psychologistic attitude towards rationality. The view that knowledge does not entail justification yields the elimination of a further psychologistic element from our epistemology. On this view, not only is truth seen as independent of the knowing subject, but knowledge itself comes to be understood without reference to the knowing subject's being justified in her own (objectively true) beliefs.

---

[9] Feldman (1981, 266) describes 'fallibilism' as the claim that 'It is possible for *S* to know that *p* even if *S* does not have logically conclusive evidence to justify believing that *p*'. Anyone who fails to believe that epistemic justification entails truth will come out as a fallibilist according to this definition. A more stringent definition of fallibilism might be provided by the following claim: 'It is possible for *S* to know that *p* even if *S*'s epistemic justification for believing that *p* is defective or nil'. Our externalist view of knowledge is clearly committed to the truth of this claim, because it does not view the possession of justification as a necessary condition for knowledge.

The justifiedness of beliefs is, I have argued, a context-dependent matter: the justifiability of a given belief typically depends on the audience the knowing subject is confronted with. Particularly relevant to the psychologism issue is the fact that epistemic justification depends on the available evidence, which may make it subject-relative (*my* being justified in believing that I have a headache does not entail that everybody is justified in believing that I have a headache, nor does it enable me to justify before a suspicious audience the claim that I do) as well as species-relative (*my* being justified in believing that I see a red spot does not entail that any alien beings confronted with the same situation would be justified in believing that they saw a red spot). Accordingly, regarding epistemic justification as unnecessary to knowledge amounts to a de-psychologization of our notion of knowledge. In other terms, in the same way as truth is thought by Frege and Husserl to be independent of what any particular subject may believe, knowledge is now thought to be independent of what any particular subject may be justified in believing. This is a significant result, because it means that whether S knows that *p* is now thought to be independent of the criteria of epistemic justification endorsed by S or by any other individual or group of individuals — independent, that is, of the way in which S or any individual or group of individuals may decide to *evaluate* the epistemic credentials of the claim that *p*[10]. The reason why I prefer to describe this view of knowledge as naturalistic rather than as purely externalist is that for all its anti-psychologistic features it does not commit its supporters to deny that *internal* epistemic justification is necessary for backing our knowledge claims.

[10] It may be helpful to notice, however, that this view of knowledge would *not* come out as anti-psychologistic on the basis of Kitcher's (1992, 59-62) definition, according to which every analysis of the 'third condition' that fails to be framed in purely logical terms will yield a 'psychologistic' epistemology. Of course labels are of little importance. My terminology presupposes that the central point of Frege's and Husserl's anti-psychologism was to reject the 'speciesism' implicit in the empiricist idea that logic and epistemology describe the actual psychological processes of the members of the human species. And surely the view that S's knowing (or failing to know) that p does not depend on what S may be (internally) justified to believe has the consequence that S's knowing (or failing to know) that p can be defined without any reference to the psychological nature of the members of S's species.

Chapter 6

Epistemology Naturalized

Having done my dutiful best to reject the claim that the hologram and clairvoyance stories embed genuinely internalist intuitions capable of refuting a naturalistic approach to knowledge, I now turn to the issue of the *rationality* of our cognitive efforts. For even if it is true that those stories do not provide unequivocal evidence for anything over and above the existence of thin epistemic justification, human beings would lack any 'control' over their epistemic access to correspondence truth if they were totally unable to tell the difference between their knowing that *p* and merely apparently knowing that *p*. In this chapter I propose to investigate how far their ability to tell the difference between their knowing that *p* and merely apparently knowing that *p* must reach if they are to stand any chance of improving their epistemic situation.

The traditional answer to this question is that human beings will not stand any chance of improving their epistemic situation unless they have access to an a priori canon of methodology — what in the last chapter I called a 'Grand Normative Theory' — which can serve as a guide for their cognitive decisions. The rejection of this claim will lead me to outline the main features of the naturalistic approach to cognitive prescriptions one must be prepared to subscribe to in order to defend the claim that if there is anything we can know, that is reality as it is 'in itself'.

If S's knowing that *p* is a real, factual connection between S and her environment and epistemic justification cannot be seen as a 'third condition' for knowledge, we cannot expect to derive any substantive methodological advice from our analysis of knowledge alone. What can count as *evidence* that S knows that *p* will be — both for

S herself and for subjects other than S — a factual issue to be decided on empirical grounds. The main point I am going to argue in this chapter is that although the prospects of successfully developing a Grand Normative Theory recommending optimal strategies for the totality of our cognitive efforts look rather gloomy, this does not prevent us from developing and exchanging a number of humbler, content-specific methodological prescriptions which can lead to dramatic improvements in our cognitive situation. Such methodological prescriptions cannot be shown to be valid in all possible worlds, but their effectiveness allows the same kind of context-relative justification that is characteristic of lower-level knowledge claims.

We saw in chapter 3 that internalist epistemologists believe that no genuine instance of knowledge can fail to be epistemically transparent to the cognitive subject to which it belongs: the very fact that S knows that *p* implies that S cannot fail to have epistemic access to the fact that she knows that *p*. As a consequence of this belief about the nature of knowledge, internalist epistemologists typically think that it must be possible to find out a priori (i.e., independently of all experience[1]) what distinguishes (true) beliefs that qualify as knowledge from (true) beliefs that don't. (This is not literally a logical consequence, but it seems natural to suppose that, if the evidence available to S in virtue of her knowing that *p* suffices to give S epistemic access to the fact that she knows that *p*, then a careful application of the method of conceptual analysis should enable one to find out a priori what it is that *in general* distinguishes true beliefs that qualify as knowledge from true beliefs that don't).

It is precisely because internalist epistemologists believe that it is possible to find out a priori what distinguishes (true) beliefs that qualify as knowledge from (true) beliefs that fail to do so, that they also think that a successful analysis of epistemic justification will enable them to dictate universally applicable rules for the best conduct of our cognitive efforts. Rational inquirers, on the other hand, are also engaged in the attempt to find out what distinguishes (true) beliefs that qualify as knowledge from (true) beliefs that don't. But they are not constrained to formulate universally applicable rules for the

[1] The Kantian characterization of a priori knowledge as knowledge that is independent of all experience is of course far from satisfactory. We saw in section 4.4 that one can find out by perceptual means that a given triangle is such that the sum of the lengths of AB and BC exceeds the length of AC. And of course one can be taught a theorem. But even if some necessary truths are sometimes known by perceptual means, there is a sense in which they could have been known (if they had been believed) independently from all experience. For an attempt to provide a precise definition of the notion of a priori knowledge, see Kitcher (1985).

best conduct of their cognitive efforts, and therefore do not hesitate to apply whatever relevant knowledge they think they have to the specific problems they are trying to solve.

My claim in this chapter will be that there is little in common between the activities of internalist epistemologists and rational inquirers. Engaging in a rational inquiry, far from presupposing access to a Grand Normative Theory, represents the only sensible way in which human beings can hope to improve their knowledge of the world: the development of cognitive prescriptions is not the business of an a priori philosophical discipline, but part and parcel of concrete scientific practice.

Section 6.1 discusses and rejects two popular arguments for a priori epistemology. Section 6.2 argues that no substantive inductive policy, and hence no Grand Normative Theory, can be vindicated on purely a priori grounds. Section 6.3 describes two ways of justifying cognitive methods a posteriori. Section 6.4 argues that cognitive methods which are reliable only in a subset of all possible worlds are all we need to improve our cognitive situation. Sections 6.5 and 6.6 discuss and reject versions of evolutionary epistemology and scientific realism which seem to me to stretch the naturalistic approach to epistemology too far. Finally, section 6.7 contrasts the naturalistic view of the assessment of cognitive methods (which involves the claim that truth is prior to methodology) to the picture arising from commitment to epistemic truth (which involves the claim that methodology is prior to truth).

6.1 A Priori and Naturalized Epistemology

There is an obvious link between the enterprise of traditional epistemologists as typified by Descartes' 'project of pure inquiry' and the belief that epistemology should be done a priori. Today it is commonly acknowledged that a false belief can be epistemically justified; and yet, if the goal of epistemic justification is taken to be the 'epistemization' of beliefs, the probabilistic relationship between epistemic justification and truth which is supposed to vindicate the rationality of our cognitive practices will have to be established *before* we set out to apply our methodological canon to the investigation of the world. In section 3.4 the internalist requirement that *all* instances of knowledge be epistemically transparent to the knowing subject to which they belong was argued to

have disruptive consequences for the viability of the project of traditional epistemology. On the other hand, there are at least two arguments which are meant to show that the only way in which human beings can develop valid cognitive prescriptions is by engaging in an a priori discipline called 'epistemology'. According to these arguments, anything short of a Grand Normative Theory developed on purely a priori grounds will fail to provide rational guidance for the improvement of our present epistemic situation. To these two arguments I now turn my attention.

The first argument for a priori epistemology rests on the impossibility of providing an uncontrovertible justification of cognitive prescriptions by appealing to a posteriori knowledge of the world. The idea is, roughly, that there must be a set of cognitive prescriptions which are independent of any assumption about what the world is like. For if all cognitive prescriptions depended on some substantive assumption about the nature of the world, then the epistemic status of all our beliefs would depend on a number of substantive truths about the world that would themselves defy all attempts at a non-circular justification.

There are more and less compelling versions of this argument, but all of them appear to presuppose a non-Kripkean view of the relationship between the metaphysical categories of necessity and contingency on the one hand, and the epistemological categories of a priori and a posteriori on the other. They presuppose, in fact, the traditional view that our knowledge of necessary truths must be independent of all experience, i.e., that it must be a priori.

To present the argument at (what I take to be) its best I need a little bit of terminology. I shall say that a cognitive prescription is 'valid' when it is conducive to our cognitive aims (whichever they are: the relativity of this notion to specific sets of cognitive aims can be ignored for our present purposes). And I shall use the terms 'general' and 'local' to describe those (valid) cognitive prescriptions which are conducive to our cognitive aims, respectively, in all possible worlds, or merely in a proper subset of them. Since most cognitive prescriptions are conditional in form, by saying that a given cognitive prescription is valid in all possible worlds (i.e., generally valid) I mean, more precisely, that it is valid in all possible worlds *in which it is applicable*. A locally valid cognitive prescription is not a prescription which is not generally applicable, but one which is valid only in a proper subset of the possible worlds in which it is applicable. Once our cognitive aims are fixed, it will be a matter of fact whether a given

cognitive prescription (or a given set of cognitive prescriptions) is conducive to their fulfilment or not. Hence it is a consequence of the traditional view of the relationship between the a priori and necessity that whether a *general* cognitive prescription is valid must be knowable (if at all) a priori. This is because a cognitive prescription could not be conducive to our cognitive aims in all possible worlds if its validity presupposed the truth of some substantive claim about the structure of the cognizer's world. Since a cognitive prescriptions which is *generally* valid must be conducive to our cognitive aims *whatever our world is like*, its validity must be knowable (if at all) a priori.

As the validity of general cognitive prescriptions must be knowable (if at all) a priori, to prove the indispensability of a priori epistemology one only needs to establish that the (supposedly uncontroversial) fact that human beings are capable of making valid cognitive prescriptions requires the existence of such generally valid cognitive prescriptions. The argument is then that if our capacity of making (locally) valid cognitive prescriptions is not to be totally accidental, we must be able to know what the actual world is like, or else we could not know that our (local) cognitive prescriptions are in fact (i.e., in our world) effective means to our cognitive ends.

Suppose that cognitive prescription P is valid only in p-worlds; then we must be able to know that our world is a p-world if we are to know that P is valid in our world. But this means that there must be a non-empty set of cognitive prescriptions which are valid both in p and not-p worlds and which are strong enough to rule out the possibility of our world being a not-p world. If these cognitive prescriptions are generally valid, they will have to be knowable a priori. If, on the contrary, their validity is restricted, say, to q-worlds, then to know that they are valid in our world we shall need a non-empty set of cognitive prescriptions which are valid both in q and not-q worlds and which are strong enough to rule out the possibility of our world being a not-q world. But we cannot be satisfied with cognitive prescriptions which are merely locally valid *ad infinitum*; if P is to be known to be valid, at some point we shall have to find a non-empty set of cognitive prescriptions which are valid *whatever the world is like*. But we have seen that those cognitive prescriptions which are valid in all possible worlds must be knowable (if at all) a priori. The capacity of (non-accidentally) making locally valid cognitive prescriptions has thus been shown to presuppose the existence of cognitive prescriptions which can be known to be valid a priori. These are the principles that have been traditionally thought to be the subject of pure, normative epistemology, which is

therefore regarded as an a priori discipline. The argument can thus be regarded as proving the indispensability of a priori epistemology from the assumption that human beings are capable of (non-accidentally) making valid cognitive prescriptions.

Kripkean objections aside, this argument seems to be valid only if 'knowing' is construed in an internalist sense — or, at any rate, if S's knowing that *p* is taken to require S's being internally epistemically justified in believing that *p*. In particular, it is true that we must be able to know what the actual world is like to be able to know that our local cognitive prescriptions are valid. But knowing what the actual world is like need not involve being epistemically justified in believing that the actual world is, say, a *p*-world. Accordingly, on a naturalistic view of knowledge, knowing that a local cognitive prescription is valid here and now does not require that we be able to rule out the possibility of our world being a not-*p* world on the basis of a set of principles which are knowable a priori (or knowable any other way for that matter).

So much for the possibility of *knowing* the validity of local cognitive prescriptions; what about the possibility of vindicating them, i.e. of *justifying* the claim that they are conducive to our cognitive aims? The fact that one can know that *p* without being justified (or being able to justify) one's belief that *p* of course does not mean that there are beliefs we are not entitled to criticize: *any* part of our system of beliefs can become the subject of a request for justification and *any* (knowledge) claim to the effect that *p* involves a commitment to provide a (context-relative) justification of it. Is it really the case that by subscribing to an externalist view of knowledge one can make sense of the activity of *making* cognitive prescriptions without taking refuge in a priori epistemology on the one hand, and without succumbing to the charge of vicious circularity on the other? The answer is, Yes. From the fact that a normative epistemology providing a set of general principles valid in all possible worlds would have to be knowable a priori it does not follow that every attempt at providing a (context-relative) vindication of the reliability of humbler, content-specific cognitive prescriptions on a posteriori grounds is going to be viciously circular because it is committed to rely on knowledge of that very empirical sort it is supposed to vindicate. Of course some precautions will have to be taken. It would be illegitimate to vindicate the reliability of some method *M* by appealing to *M* itself or to beliefs which cannot be justified other than by appealing to *M*. But as long as one takes care to avoid these pitfalls, there is no obvious reason why every attempt to vindicate cognitive prescriptions

as valid in a limited subset of all possible worlds should be viciously circular (in section 6.3 I will describe in greater detail two valid ways of vindicating cognitive prescriptions on a posteriori grounds).

The second argument for a priori epistemology appeals to the normative character of cognitive prescriptions. The claim is that these cannot be based on psychological or sociological knowledge of the cognitive methods actually followed by this or that group of inquirers, but only on a priori knowledge of the methods all rational inquirers *ought* to follow. This second argument is a straightforward application of the currently unfashionable view that 'ought' cannot be derived from 'is' (that is, of the view that the 'naturalistic fallacy' is indeed a fallacy). The point is supposed to be that if cognitive prescriptions are to describe the methods rational inquirers *ought* to follow, then investigating what methods *are* followed by this or that group of inquirers will be wholly irrelevant to the success of the project. Valid cognitive prescriptions can thus be provided only by an a priori discipline to be identified with normative epistemology.

As it stands, this argument is of course a *non sequitur*. For it seems plausible to think that the cognitive methods rational inquirers ought to follow are those methods that are maximally effective in fulfilling their cognitive (i.e., inherently truth-linked) aims. But then, even if it is true that investigating what methods *are* followed by this or that group of inquirers will not help one to identify those methods rational inquirers *ought* to follow, it is obviously false that investigating what methods *are* most effective in fulfilling the relevant cognitive aims will be irrelevant to the development of cognitive prescriptions. However, there is a way of formulating this second argument for a priori epistemology that makes it sound more plausible. This is by presupposing a 'categorical', as opposed to a merely 'hypothetical' (i.e., instrumental), view of rationality.

Categorical theories of rationality can be found, for example, in Strawson (1952, 256-257; 261-262) and Chisholm (1989). According to such theories, being rational is not, in general, a matter of using effective means to desired goals. Some philosophers (most notably, Aristotle) write as if being rational was an essential property of human beings; be that as it may, the central point of these theories is that human beings can know a priori several conceptual truths about rationality. This means that there are a number of criteria that are constitutive of human rationality independently of any consideration of their effectiveness or reliability in producing truth-acquisition and error-

avoidance. The criteria (or some of the criteria) for evaluating the epistemic justification of our beliefs are not adopted because they are believed to be effective means to our cognitive aims: they are adopted for the pure and simple fact that they unfold the meaning of our concepts of 'rationality' and 'epistemic justification'. This applies, according to Strawson, to the method of induction:

> the rationality of induction, unlike its 'successfulness', is not a fact about the constitution of the world. It is a matter of what we mean by the word 'rational'.
> (Strawson 1952, 261)

On this view of rationality, factual questions concerning the reliability of cognitive methods are indeed irrelevant to normative epistemology, provided this is understood as the project of identifying 'rational', and not 'successful', cognitive methods.

Now, if the whole point of epistemology is to investigate what we mean by the word 'rational', it is clear that epistemological knowledge will be a priori in the very sense conceptual analysis is commonly thought to be. But this means that the guidance epistemological knowledge can be expected to provide for our cognitive efforts will be a very poor thing. It is small wonder, then, that 'categorical' theories of rationality are nowadays increasingly unpopular: if our criteria of epistemic justification are not even *believed* to be effective means to our cognitive aims, why should we care about them in the first place? Why should we bother to unfold a concept of rationality which is *ex hypothesis* completely divorced from any consideration of cognitive success? If our cognitive methods are not selected because they are thought to be effective means to our cognitive aims, but merely because they conform to our idea of 'rationality', it is very hard to see why one ought to take pains to be 'rational' in the first place. If the reason why epistemology must be an a priori discipline is that it must be normative, appealing to a categorical conception of rationality in order to ground its prescriptions appears to be the wrong way to support the claim that epistemology can provide helpful indications as to how we can foster the fulfilment of our cognitive aims.

6.2 The Vindication of Inductive Policies

That the project of vindicating a Grand Normative Theory on purely a priori grounds (and without subscribing to a categorical theory of rationality) is doomed to failure is strongly indicated by the impossibility of identifying substantive inductive policies that will be preferable to others *whatever the world is like*. This point can be brought out by examining Hans Reichenbach's classical attempt to vindicate the 'straight rule' by showing that its use is going to succeed *whatever the world is like* (the following discussion relies on Friedman 1985, 149-153).

Reichenbach (1949) claims that the straight rule — if you have observed that a certain kind of event has appeared with a relative frequency f^n in a given sequence of events, predict that its relative frequency for $n \to \infty$ will be equal to f^n — is vindicated *whatever the world is like* because it is a consequence of the definition of 'limit' that in the long run use of the straight rule will converge on the limiting relative frequency, *if it exists at all*. This is because if there is a relative frequency for $n \to \infty$ (call it p), then for any arbitrarily small positive number ε, there will be a number N such that p will lie within ε of f^n for all $n \geq N$.

Apart from the obvious fact that whether the limiting relative frequency of a certain kind of event exists at all does depend on what the world is like (and therefore use of the straight rule will be successful only in those worlds in which there *is* such a limiting frequency), the main problem with Reichenbach's a priori vindication of the straight rule is summarized by Keynes's reminder that 'in the long run we are all dead'. The only sequences that provide a ground for inferring that p lies within ε of f^n are those in which $n \geq N$. But the definition of limit only tells us that there *is* some N such that p lies within ε of f^n for all $n \geq N$. It does not tell us what this N actually is. In particular, it does not tell us 'whether the inferences we actually do make, or even whether the inferences that are physically possible for us to make, occur before or after the point of convergence' (Friedman 1985, 151). This predicament is made even more serious by the fact that Reichenbach's a priori vindication of the straight rule cannot tell us what the relationship between f^n and p is with $n < N$.

But this are not the only difficulties Reichenbach's approach has to face. Reichenbach himself recognized that the straight rule is just a special case in a general class of methods, all members of which converge, in the long run, on p (if it exists).

This is because if c_1, c_2, c_3, ... is any sequence of positive numbers such that $\lim_{n \to \infty} c_n$ is equal to 0, then

$$\lim_{n \to \infty} (f^n + c_n) = \lim_{n \to \infty} f^n.$$

This means that all the methods that predict that p lies within ε of $f^n + c_n$ will yield the same result 'in the long run'. The straight rule is then just the special method that results from assuming that $c_n = 0$. But how can this assumption be vindicated on purely a priori grounds if all these methods can be shown to converge on the same p? How can we tell that our world is such that the straight rule will lead to true conclusions sooner then the other methods? 'A priori, there is no way of knowing which value, including the value $c_n = 0$, will lead to convergence soonest. A priori, all values of c_n are equally risky' (Friedman 1985, 152).

A further problem for the project of vindicating the straight rule a priori arises from Goodman's (1973, 72-81) well-known point that many different and mutually incompatible inductive policies can be regarded as applications of the straight rule, provided suitable ('grue-like') predicates are employed in the description of the properties of the events in the observed sample.

The severity of these problems suggests that Reichenbach's project of vindicating our inductive practices as rational *in all possible worlds* is doomed to failure. For being told that all those inferences that can be formalized as applications of a member of the above mentioned class of inductive methods will eventually converge on the true value of p will not help us to identify those *substantive* inductive policies we had better follow if we want to maximize our chances of cognitive success *here and now*. We need cognitive methods which will help us to achieve our cognitive aims before we are all dead; cognitive methods which are conducive to our cognitive methods nobody knows where and when will not do. But if it is true, as — *pace* Popper — it is true, that inductive practices play a central role in our cognitive undertakings, the failure of Reichenbach's project suggests that an epistemology developed on purely a priori grounds will have very little to say about how we ought to proceed in order to improve our present stock of beliefs.

I draw this general conclusion because, although Reichenbach's approach is of course just one of many attempts to provide an a priori vindication of our inductive

practices without appealing to a categorical view of rationality, the problems haunting his approach are absolutely general (various versions of the 'Dutch Book' argument have been claimed to provide something very close to an a priori vindication of 'subjective' or 'personalist' Bayesian methodology, but again, the reliance of this methodology on 'long run' considerations for the 'washing-out' of prior probabilities prevents this family of arguments from representing a vindication of *substantive* inductive policies). A Grand Normative Theory providing cognitive guidance *whatever the world is like*, far from representing the precondition for the development of any cognitive prescription at all, appears to collapse as hopelessly empty. It looks as if we had better reconcile ourselves to the idea that most of our cognitive methods (and prescriptions) are such that they cannot be vindicated as valid in all possible worlds, but only in a small subset of them, the actual world hopefully included.

6.3 Calibration against Standards and Theoretical Justification of Cognitive Methods

That the vast majority of our cognitive methods cannot be vindicated as valid in all possible worlds but only in a small subset of them may appear as a problem to someone who believes that the aim of epistemic justification is to turn true beliefs into knowledge. This is because if no substantive cognitive method can be vindicated as successful in *all* possible worlds, no substantive belief about the *actual* world can legitimately be regarded as a genuine instance of knowledge. But the fact that the majority of our cognitive methods cannot be vindicated in all possible worlds is emphatically not a problem for someone who believes that the aim of epistemic justification is to affect other people's and one's own beliefs in order to maximize their mutual agreement and their agreement with reality. For *this* project does not require that we reconstruct our stock of beliefs from scratch, but merely that we start from our present epistemic situation — which may well involve the possession of some actual knowledge of the world — and do everything in our power to improve our present stock of beliefs.

We need some knowledge of the actual world to tell which cognitive methods will be most effective in *our* world. But this is not a problem if knowledge is understood, as I have argued it should be, as a real, factual relationship between a subject and her

environment. Knowledge that is not transparent to the cognitive subject to which it belongs can be used to vindicate cognitive methods which in turn may lead to dramatic improvements in the subject's own cognitive situation. Perceptual knowledge, for example, is unlikely to have been epistemically transparent to our ancestors (for that matter, it is unlikely to be epistemically transparent to most contemporary psychologically unsophisticated adults). And yet reliance on perceptual knowledge has apparently led human beings to many substantive discoveries about the world as well as to the development of many content-specific cognitive methods. How are such achievements possible?

Jardine (1986, 95-111) identifies two ways in which content-specific cognitive methods can be vindicated on a posteriori grounds: (i) by calibration against standards, and (ii) by theoretical justification. These procedures can be applied to assess the reliability of the evidence provided by measuring instruments (including the perceptual apparatus of human beings), but they can also be applied to assess the reliability of higher-level methods of theory-evaluation.

Calibration against standards is perhaps the most rudimentary procedure for the evaluation and selection of cognitive methods one can think of. Consider the issue of the reliability of visual perception. In this case, independent knowledge of the physical properties of certain objects may enable one to discover under what circumstances the evidence provided by visual perception is apt to be affected by various kinds of optical 'illusions'. Or think of the voltmeter case discussed in section 4.3: that sort of instrument can be calibrated by measuring the independently known values of the voltage drop across different points of a familiar electric circuit. What is required for the success of a procedure of calibration is, in general, the possibility to tell on independent grounds whether, say, the stick immersed in the water is 'really' bent, or whether the voltage drop across the resistor is 'really' 5 volts. Instruments are calibrated by comparing the results of their application to one's independent knowledge of what the 'right' results should be.

Now, the fact that we succeed in calibrating various sorts of detecting processes on the basis of independent knowledge of what the 'right' results should be explains why I insisted that any adequate view of knowledge ought to make room for the possibility that S know that she knows that p. For it is clear that the upshot of (successful) calibration processes is precisely to enable one to *tell* whether one knows that the stick

immersed in water is bent when it looks bent, or whether one knows that the voltage drop in the circuit is 12 volts when the pointer of the voltmeter comes to rest on 12. One cannot engage in activities of method-evaluation without learning to *tell* whether one knows the sorts of things those methods are supposed to give one epistemic access to (recall the characterization of 'S knows that she knows that *p*' given in section 4.4).

Our examples so far have concerned the reliability of low-level cognitive methods involving the use of visual perception and simple measuring instruments. The way in which calibration against standards can be used to evaluate the reliability of higher-level cognitive methods is formally equivalent:

> Suppose that *T* conflicts with *T'* and that there are grounds independent of the reliability of *M* for holding *T* to be true, or a better approximation to the truth than *T'*; then if *M* applied to the conflict between *T* and *T'* adjudicates in favour of *T*, the reliability of *M* is confirmed.
>
> (Jardine 1986, 103)

The viability of the enterprise depends, once again, on the availability of independent evidence for the truth (or falsity) of some of the results of the application of the method under scrutiny. And success will involve, just as before, the acquisition of the ability to *tell* whether the application of that method enables one to know the sort of things it is supposed to give one epistemic access to.

The vindication of cognitive methods through *theoretical justification* requires something more than the availability of independent evidence for the truth (or falsity) of some of the results of the application of the method under scrutiny. It requires the possession of a body of independent knowledge that can now have only an indirect bearing on the phenomena falling under the methods at issue. We can again refer to the example of visual perception. When we assess the reliability of visual perception by calibration against standards, we treat the visual process as a black-box, and we do not try to explain the reliability (or lack thereof) of its output. Needless to say, there is apparently no other way in which our ancestors could have assessed the reliability of visual perception. But *we* are in possession of independent optical and physiological information that is obviously relevant to the problem. The genesis of this information is not independent from the operation of our perceptual apparatus, but what really matters is that it is not directly *about* our perceptual apparatus. Therefore we can appeal to this information to devise a theoretical justification of the reliability of visual perception in

the circumstances in which it is reliable, and a theoretical explanation of its delusory character in the circumstance in which it is not.

At a higher level, Jardine discusses the methodological rule that in evaluating hypotheses about the impact on plant distributions of changes in the disposition of land masses one should attach little importance to past or present distributions of spore-bearing plants or of lighter seeded flowering plants. This particular methodological rule can be justified a posteriori on the basis of direct evidence for the aerial transport of spores and of uncontroversial theoretical knowledge about the transport of bodies in turbulent airstreams such as the trade winds (Jardine 1986, 97).

But of course there will be theoretical dilemmas for the solution of which no theoretically justifiable cognitive rule is forthcoming. This is because the theoretical justification of cognitive methods requires a substantial background of uncontroversial theory. When such a substantial background is unavailable, one will perhaps have to take refuge in that kind of very general methodological rules that can be supported (if at all) only by calibration against standards.

I am thinking of the sort of rules that make up Lakatos's (1978) methodology of scientific research programmes, which are theoretically unjustifiable and prescriptively almost empty, and yet appear to make sense of a considerable amount of (what we take to be) the best science of the past. Lakatos's (1978, 102-138) 'historical' criterion for the evaluation of competing methodologies — prefer the methodology that provides the best 'rational reconstruction' (i.e., the reconstruction that maximizes the rationality) of what the scientific élite considers the best science of the past — yields in effect a high-level instance of calibration against standards[2]. The standards involved are uncontroversial (at least among the members of the scientific élite) judgements about the worth of certain scientific achievements. The success of a given set of methodological rules in providing a rational reconstruction of those achievement provides a vindication of the reliability of those rules, in so far as they are shown to have been effective means to (what is now regarded) as cognitive success. Of course the impossibility of performing actual *tests* of the reliability of methodological rules of the generality of those

[2] Jardine (1986, 197) rejects this conclusion on the ground that 'progress and rationality are rarely linked by Lakatos with approximation to the truth or increase in truth content'. But this reading of Lakatos's work, which is fully elaborated in Hacking (1983, 112-128), fails to do justice to Lakatos's (1978, 159-166) 'plea' for an inductive principle relating scientific progress to increasing verisimilitude.

recommended by Lakatos's methodology of scientific research programmes heavily affects the cogency of this kind of vindication.

A successful defence of the epistemic accessibility of correspondence truth need not provide, however, a proof that human beings can develop effective cognitive methods for the resolution of *all* conceivable scientific dilemmas. The case against correspondence truth is supposed to be that it is bound to make even the most trivial truths about the world epistemically inaccessible to our cognitive efforts (see section 2.1 above), not that it fails to provide a wholly general solution to the problem of the underdetermination of theory by empirical evidence. The fact that there may be theoretical issues the empirical underdetermination of which is entailed by the very physical theories we currently accept is no refutation of the claim that human beings can develop effective cognitive methods for improving their knowledge of reality as it is 'in itself'. Even though I argued that metaphysical realism (the claim that the world would exist and retain its structural properties even if it were the case that there were no cognizers around) does not entail semantic realism (the claim that certain statements about the world could be true even if no human being should ever be in a position to verify them), it is small wonder that significant portions of reality may ultimately escape our grasp.

6.4 Reliability in a Subset of All Possible Worlds

With the possible exception of strictly formal rules such as, 'One ought to eliminate logical contradictions from one's system of beliefs', it appears that most methodological principles make in fact substantive assumptions about what the world we happen to live in is like. That means, as Larry Laudan puts it, that

> the cogency of any methodological principle is, at least in part, hostage to the vicissitudes of our future interactions with the natural world. But that is just another way of saying that methodologies and theories of knowledge are precisely that, theories. Specifically, our methodological rules represent our best guesses about how to put questions to nature and about how to evaluate nature's responses. Like any theory, they are in principle defeasible. And like most theories, they get modified through the course of time.
>
> (Laudan 1989, 374)

The history of science (of post-Galilean science!) is a mine of examples of methodological changes affecting the criteria of epistemic justification accepted by the scientific community. Saying this does not commit me to deny the possibility that there might be a few very general cognitive methods endorsed by post-Galilean science as a whole. But it is hardly deniable that there have been in fact several *local* methodological shifts which have progressively modified our view of what can be counted as an epistemically justified scientific belief. Here are a few examples.

Non-deterministic theories failing to ascribe sharp values to some of the quantities involved in the description of a physical system would have been methodologically unacceptable for 18th and 19th century physics. Yet they are now regarded as fully acceptable by the style of reasoning of present-day physics, since they play a major role in contemporary quantum theory.

A methodological rule which came to be adopted only relatively recently is that clinical trials in medical science are to be performed 'double blind'. Its adoption was the consequence of the recognition that patients are subject to the placebo effect and can be affected by the therapeutic expectations of people administering drug tests. The adoption of this rule clearly makes some difference to what medical beliefs will be regarded as epistemically justified by present-day physicians.

A content-specific rule that is closely connected to evolutionary theory is that we should not treat evolved characteristics of a given type of organism as being necessarily conducive to its survival in the relevant environment, for there could be another adaptive characteristic of the organism of which the former is merely a by-product.

All these rules were adopted as a consequence of the discovery of some substantive truth about what the actual world is like. This does not mean that a non-circular justification of their adoption is impossible, although it does mean that they cannot be proved to be successful in all possible worlds, but only in a limited subset of them. This, however, is as it should be. Reliability across a subset of all possible worlds is, if the actual world belongs to that subset, everything one needs in order to improve one's present epistemic situation. The only advantage afforded by (provable) reliability across all possible worlds would be the additional peace of mind provided by the availability of a refutation of scepticism. However, the fact that S's knowing that p need not be taken to imply S's ability to rule out all sorts of irrelevant alternatives to p suggests that S's knowing that M is a reliable cognitive method in the actual world need

not be taken to imply S's ability to rule out any sort of irrelevant alternatives to the actual world. If S knows that the actual world is *p* and has independent evidence that method *M* is successful in all *p*-worlds, then the mere fact that it is possible to conceive some evidentially indistinguishable alternative to the actual world in which *p* would be false does not imply that *M* cannot be known to be a successful method in the actual world.

We saw in previous chapters that reference to the reliability of cognitive methods plays a crucial role in those analyses of epistemic justification [knowledge] according to which, roughly, a true belief is epistemically justified [an instance of knowledge] iff it has been produced by a reliable cognitive process. The issue of the *range* of possible worlds in which the relevant cognitive method must be reliable for the truth conditions of 'S justifiedly believes [knows] that *p*' to be met is of course particularly pressing for reliabilist analyses of epistemic justification [knowledge]. As far as this issue is concerned, reliability across all possible worlds is evidently out of question, but reliability only in the actual world seems equally implausible. Dice-throwing in the moonlight, for example, might be described as a perfectly reliable cognitive method in a world *w* in which all beliefs arrived at by dice-throwing in the moonlight happened to be unfailingly true. However, if the reliability of *M* were just the effect of an extremely improbable cosmic coincidence and not the consequence of some underlying structural property of *w*, it would be extremely counterintuitive to regard beliefs arrived at by *M* as epistemically justified [as genuine instances of knowledge]. Therefore, when the reliability of cognitive methods is discussed with reference to the analysis of epistemic justification [knowledge], it seems sensible to require, as Alvin Goldman (1986, 107-109) does, that the relevant cognitive method be reliable in all those possible worlds which are sufficiently similar to what the actual world is (rightly or wrongly) believed to be like.

However, the issue of this chapter is neither the analysis of epistemic justification nor the analysis of knowledge, but merely the vindication of cognitive methods. As far as the vindication of cognitive methods is concerned, there is no *prima facie* reason to require reliability in more than one (i.e., the actual) world. Vindicating a cognitive method as reliable in the actual world is a valuable result in itself, independently of the existence of a larger class of possible worlds to which the method can be safely applied. Still, being able to vindicate a method as cognitively reliable in a whole class of worlds

is in some respects preferable to being able to vindicate it as cognitively reliable just in the actual world. For we aspire to know *why* the methods that are reliable in the actual world are so; we aspire to single out the features of the actual world that make such methods as reliable as they are. But vindicating cognitive methods by theoretical justification necessarily means vindicating them as reliable in all those worlds that share the relevant features that make it reliable in the actual one. Of course, cognitive methods do not become any more reliable for being supplemented with an explanation of their reliability; but surely our confidence in the success of their operation is increased by knowing why they are successful in the actual world. Moreover, although methods that are vindicated by calibration against standards may very well be reliable in worlds other than the actual, we cannot know what these worlds are like until we learn *why* the methods in question are reliable in the actual world. This is worth noticing because if a method which we thought to be reliable turns out not to be so, this finding will give us much more (negative) information about the world if the grounds of our belief in its reliability were 'theoretical' than if they were merely 'observational'.

It may be of some interest to notice that there are methods which cannot even in principle be vindicated as cognitively successful other than by calibration against standards. Dice-throwing in the moonlight is a good example. If its reliability is merely a matter of cosmic coincidence, the worlds in which dice-throwing in the moonlight is a reliable cognitive method will be just those worlds in which the beliefs arrived at by dice-throwing in the moonlight are unfailingly true. There may be a number of such worlds, but they will not share any general feature capable of explaining the reliability of dice-throwing in the moonlight. This being a brute fact, it will not admit of other kind of vindication than mere calibration against standards. We can conclude, then, that even if reliability in the actual world is not to be underrated, vindicating a cognitive method by theoretical justification requires more than that, in so far as it involves the subsumption of the actual world under the class of all worlds that share the general features invoked in the justification of the reliability of the method in question.

6.5 On 'Finding Encouragement in Darwin'

Reliability in a subset of all possible worlds is, I have argued, all one can legitimately claim for human cognitive methods. There are however two types of realist arguments which are designed to vindicate a more optimistic view of our cognitive situation that I want to discuss (and reject) in this section and the next one. These arguments can be found in the writings of 'evolutionary' epistemologists (section 6.5) and upholders of that brand of 'scientific' realism commonly associated with the name of Richard Boyd (section 6.6).

The label 'evolutionary epistemology' covers different philosophical positions, ranging from the claim that the structure of human learning parallels in some interesting respects the structure of natural evolution as a blind-variation-and-selective-retention process (Campbell 1974) to the claim that our present picture of the world, being the result of the application of cognitive faculties which were selected for their reproductive success, cannot be too far from the truth. It is this latter claim that I want to discuss and reject as unwarranted in this section.

The argument that our present picture of the world cannot be substantially misguided because if it were we would not be around asking all sorts of philosophical and epistemological questions is, as an argument from the theory of evolution, obviously invalid. For the fact that our cognitive faculties were selected for their contribution to the reproductive success of the human species in a given environment does not entail that those faculties will have any tendency to yield acceptance of true theories about the world when applied to the solution of problems which are completely different to those they were originally selected to cope with. The fact that natural selection may have endowed us with the ability to acquire reliable perceptual information about our immediate surroundings does not mean that the application of our perceptual apparatus to problems in the provinces of cosmology or quantum field theory will produce any cognitive success at all.

Still, one could think that if the theory of evolution can be taken to support the view that natural selection endowed us with the ability to acquire reliable perceptual information about our immediate surroundings, one will have some reason to believe that the epistemic situation from which human beings originally started their cognitive adventure is unlikely to have been so bad as to prevent any chance of cognitive

improvement. For example, it could be argued that the ways our ancestors classified incoming stimuli must have corresponded to objective regularities in their environment, or else they would soon have been eliminated by natural selection. This kind of argument will not show that our present picture of the world cannot be very far from the truth, but one can be tempted to use it as a rejoinder against the sceptical claim that our original epistemic situation might have been so bad as to preclude any chance of improvement.

This weaker version of the argument will not do either. As many writers have recently pointed out, truth has in general little to do with either pragmatic utility or reproductive success. False theories may be pragmatically more useful than true theories, provided they are simpler to apply and give approximately correct results in those situations that really matter to our goals. The adaptive value of a cognitively infallible method which is too slow for the purposes of action will surely be outstripped by that of a quicker method which gets many cases wrong but regularly prompts an appropriate behaviour in those cases which are really important for the survival of the species. Think of mechanisms for telling the presence of predators, and compare a cognitively infallible but slow mechanism with one producing occasional false alarms but quick enough to prompt a timely reaction on the part of the potential victim when a quick decision is really important for survival. It is clearly this latter kind of mechanism that has greater adaptive value and that is more likely to have been selected by natural evolution. As Stich (1985, 258) sums it up, it is often the case that 'it is more adaptive to be safe than sorry'.

There is however a weaker argument from the theory of evolution that, far from refuting sceptical claims on the nature of our epistemic situation (which I have already said it is not my intention to refute), can nevertheless provide some independent reason for dismissing the sceptic's challenge as irrelevant for our cognitive purposes. The argument is that the impossibility of meeting the sceptic's challenge is exactly what we would expect if we had a reasonable degree of epistemic access to our world *and* the theory of evolution were true.

We have seen that there is no compelling evidence that the discriminatory processes natural selection originally endowed human beings with must be cognitively reliable. But there is no compelling evidence for the claim that those processes were so unreliable as to prevent any possibility of cognitive improvement either. If *some* of those processes were sufficiently reliable to enable human beings to improve their cognitive

situation by developing new and more effective cognitive methods, then it may well be the case that we now have a certain amount of genuine knowledge of the world. And if the theory of evolution is correct, the fact that we can possess a certain amount of genuine knowledge about the world and yet be incapable of meeting the sceptic's challenge will be hardly surprising. In fact, it is exactly what we would expect given that the discriminatory processes evolution originally endowed human beings with were selected because of their adaptive value. If the theory of evolution is right, it is hardly surprising that no 'certifying' procedure was built into those processes. For it is clear from what has been said before that quick discriminatory mechanisms that are uncapable of eliminating all relevant alternatives to a given state of affairs but can be trusted to trigger a timely response whenever it is required are adaptively much more effective than discriminatory mechanisms that are capable of eliminating a larger set of (relevant or irrelevant) alternatives, but in a significantly longer time.

Now, note that the discriminatory processes evolution originally endowed human beings with were the immediate predecessors of the perceptual processes that give us that 'basic' access to reality that constitues the (revisable) starting-point of our empirical and scientific knowledge of the world. Therefore, if later on human beings came to apply the information provided by their perceptual apparatus to the solution of increasingly abstract and speculative problems, such as counting how many chocolates are left in the box or measuring the value of Planck's constant, there is no reason why natural selection should have enowed them with the capacity of reassuring themselves of the good outcome of *this* undertaking. It is small surprise, then, that any attempt to meet the sceptic's challenge by *proving* that we know how many chocolates are in the box or the true value of Planck's constant will be frustrated by the lack of certifying procedures capable of eliminating irrelevant alternatives to perceptually known states of affairs in the first place. But then, if the impossibility of meeting the sceptic's challenge is exactly what we would expect if the theory of evolution were true and we had a reasonable degree of epistemic access to our world, one will be able to provide a 'naturalistic' account of the intractability of sceptical questions as well as a rationale for the decision of dismissing them as irrelevant to any substantive issue about our knowledge of reality.

I am not claiming that the availability of a naturalistic account of the intractability of sceptical questions will provide anything like a refutation of scepticism. For of course a naturalistic account of the intractability of sceptical questions is only

going to work if our world is not a 'sceptical' world and if the theory of evolution is approximately true. However, while the sceptic's alternatives to (what we take to be) the actual world are typically constructed in such a way that there can be no independent evidence in their favour, the naturalistic account of the intractability of sceptical questions in the actual world simply *falls out* from a theory (the theory of evolution) which was not designed to address epistemological issues in the first place (the intractability of sceptical questions does *not* follow merely from the assumption that the actual world is not a 'sceptical' world!). Therefore the availability of a naturalistic account of the intractability of sceptical questions can be seen as providing an *independent* reason for refusing to play the sceptic's game (which seems to me the wisest course of action for a realist to take). Such an independent reason will help to reject the suggestion that the refusal to play the sceptic's game is an *ad hoc* manoeuvre deliberately designed to escape the sceptical predicament.

This is, I think, the only, admittedly limited, 'encouragement' (Quine 1969, 126) naturalistically oriented epistemologists can realistically hope to find in Darwin.

6.6 Boyd's Abductive Argument for Scientific Realism

A different argumentative strategy designed to show that our present picture of the world cannot be too far from the truth can be found in Boyd's (1973) much discussed argument for scientific realism. This can be described as an abductive inference to the only possible explanation of the reliability of experimental method.

Boyd's overall approach appears to fit in very well with the naturalistic view of the vindication of cognitive methods that I have been advocating in previous sections of this chapter. According to Boyd, the reliability of our cognitive methods rests 'upon the logically, epistemically and historically contingent emergence of suitably approximately true theories' (Boyd 1983, 71). The principles of inductive inference cannot be defended a priori: we can confirm our scientific generalizations only because our predicates happen to be 'projectible' in Goodman's sense, i.e., because they happen to sort out the right features of the world (they 'cut the world at its joints'). But things could have been different, and the success of the scientific method is a merely contingent fact.

So far, so good. However, Boyd also believes that he can show that the contingent fact that the scientific method is instrumentally successful provides compelling evidence that the picture of the world provided by 'mature' science cannot be too far from the truth. And it is precisely because his argument, unlike the arguments of other upholders of scientific realism, is not merely designed to show that the theoretical claims of the scientific theories that exhibit certain features must be (approximately) true, but that the overall reliability of the scientific method involves the (approximate) truth of our present picture of the world, that I am going to discuss it extensively. For I do not want to leave the reader with the impression that the claim that we may assess the reliability of our cognitive methods and, by so doing, improve our epistemic situation involves commitment to the stronger thesis that our present picture of the world *as a whole* cannot but be (approximately) true.

The instrumental reliability of experimental method is the *fact* that Boyd's abductive argument is meant to explain. This reliability is to be construed in the minimal sense that application of the experimental method 'contributes to the likelihood that the *observational* consequences of accepted scientific theories will be (at least approximately) true' (Boyd 1973, 3, emphasis mine). Boyd points out that the instrumental reliability of the experimental method is granted even by those anti-realist philosophers that view science as exclusively concerned with 'saving the phenomena' (and even if he is commonly described as subscribing to correspondence truth, his abductive argument for scientific realism neither presupposes nor establishes any particular claim about the nature of truth; which gives me the opportunity of emphasizing that the content of this section, as well as that of the previous one, has no direct bearing on issues in the *theory* of truth).

Boyd's argument moves from the observation that currently accepted theoretical claims play a crucial role in the assessment of the experimental evidence relevant to the acceptance of a proposed theory. In particular, Boyd (1985) lists three categories of questions affecting the extent to which a proposed theory T is confirmed by the evidence E: (i) projectibility, (ii) experimental controls and experimental artifacts, and (iii) sampling. His claim is that the way we address these questions, the solution of which is crucial to the assessment of the experimental evidence relevant to the acceptance of T, is heavily theory-dependent. But if currently accepted theoretical claims play a crucial role in the operation of our methodology, then the acknowledged instrumental reliability

of our methodology must involve the approximate truth of those claims, or else the tendency of our methodology to yield acceptance of observationally true theories would be totally inexplicable.

An example may help to illuminate Boyd's argument. Take the following methodological principle:

> (P) a proposed theory *T* must be experimentally tested under situations representative of those in which, in the light of collateral information, it is most likely that *T* will fail, if it's going to fail at all.
>
> (Boyd 1973, 10)

Now consider the experimental testing of a theory *L* which specifies that some antibiotic *A* brings about lethal effects on bacterial species in some class *C* through a certain chemical mechanism *M* which causes the dissolution of the cell walls of those bacteria. Available theoretical information has it (i) that a drug similar to *A* affects those bacteria to which it is fatal not by dissolving cell walls, but by interfering with the development of new cell walls after mitosis, and (ii) that certain bacteria in *C* are particularly prone to mutations affecting the structure of the cell walls. Principle P tells us that such collateral information is relevant to the experimental testing of *L* because it suggests 'under what circumstances the causal claims made by the theory might plausibly go wrong' (Boyd 1973, 11) and hence which sorts of experiments are crucial to establishing the empirical adequacy of *L*. In particular, (i) suggests that it is crucial to test the predictions of *L* 'under circumstances involving a time much smaller than that required for the typical bacterial cell of the sort in question to divide, together with a large dosage which *L* predicts will be fatal to most bacteria in this time interval' (Boyd 1973, 11); and (ii) suggests that it is crucial to test the predictions of *L* under conditions of low dosage and over time intervals so long as to make mutations in the structure of the cell walls likely to occur.

Now, Boyd's claim is that since the application of principle P relies on 'collateral information' about the system under scrutiny, the only possible explanation of the instrumental reliability of P will be that that collateral information is approximately true. But the collateral information required by the instrumental reliability of P will typically be *theoretical* information about unobservable causal mechanisms, and from this Boyd concludes that the uncontroversial fact that the experimental method is *instrumentally*

reliable shows that 'scientific realism' ('realism-without-modality': see section 1.1 above) is true.

It should be noted that Boyd's abductive argument for the truth of scientific realism operates at a higher level than the usual 'no-miracle' argument from the predictive success of theories in 'mature science'. Its *explanandum* is not the predictive success of scientific *theories*, but the instrumental reliability of scientific *method*. This fact might be thought to immunize Boyd's argument against Laudan's (1981) 'pessimistic meta-induction' (the outcome of scientific revolutions shows that the stunning predictive success of past scientific theories is not to be explained by their approximate truth; why should the stunning predictive success of *present* scientific theories be explained by *their* approximate truth?). After all, one might say, Boyd's argument is not purported to establish any direct link between the predictive success of scientific theories and their (approximate) truth: it is only purported to establish a link between the instrumental reliability of scientific method and the (approximate) truth of the collateral information required for its operation. But in fact, by postulating a link between the instrumental reliability of scientific method and the (approximate) truth of the collateral information required for its operation, Boyd's argument does postulate a link between the predictive success of certain scientific theories — which is what the instrumental reliability of scientific method boils down to, if by 'scientific' theories one means those theories the acceptance of which is warranted by the 'scientific' method — and the (approximate) truth of *other* scientific theories. Indeed, Boyd's argument can be fairly seen as an attempt to support the (approximate) truth of those theories the predictive success of which is attested by the particular role they play in the successful selection of other empirically adequate theories.

Compared to some sloppy formulations of the 'no-miracle' argument, this methodological focus provides a more stringent characterization of the kind of 'predictive success' that is required for running the abduction to the truth of scientific realism. For the sort of predictive success that is required by Boyd's argument has to do with the testing of *new* theories, and therefore, presumably, with kinds of phenomena the behaviour of which cannot have been 'written into' the original formulation of the theory. The predictive success required by Boyd's argument involves, in other words, the confirmation of *novel* predictions falling out from the original theory when it is applied to the testing of new theoretical hypotheses. Even though this represents an

undeniable progress as compared to alternative formulations of the 'no-miracle' argument, Boyd's abduction is itself far from flawless.

First of all, one can develop a pessimistic induction at the meta-meta-level of Boyd's own argument. *How* reliable is the experimental method? Is it really true that it can be trusted to yield acceptance of observationally true theories? It seems to me that our confidence in the instrumental reliability of the experimental method cannot be without qualifications. One does not have to espouse Karl Popper's denial of the predictive significance of the corroboration of scientific theories or Nancy Cartwright's scepticism about the exportability of well-established physical laws outside the walls of our laboratories to realize that the instrumental reliability of the experimental method is far from faultless.

Even if it is true that the history of science seems to display an essential cumulativity at the observational level, this appearance is often obtained by restricting the domain of older theories when they are replaced by more precise and more general ones. In many cases, the cumulativity does not apply to the observational level *tout-court*, but only to those kinds of observations that were actually carried out before the emergence of the new theory. For example, Newtonian mechanics purported to be about *all* material objects moving with *any* possible velocity, and not merely about macroscopic objects moving with velocities small compared to that of light. However, the latter is the only domain in which classical mechanics is now believed to be empirically adequate: outside this domain, there is compelling experimental evidence that it is not so. But this means that, if Newtonian mechanics is to count as an application of the supposedly reliable 'experimental method', the instrumental reliability of that method must allow for the possibility that even our best theories may turn out to be empirically adequate *only within an appropriately restricted domain* — a domain, moreover, which we shall typically ignore *how* large is going to be. Under such circumstances, one may start to suspect that, if there is anything that the (partial) instrumental reliability of the 'experimental method' is going to tell us about the collateral information on which it relies, this will not necessarily be that that information is (approximately) true. If the application of the experimental method has so far yielded acceptance of theories which are empirically adequate only within an (unpredictably restricted) partition of their intended domain, why should we believe that the collateral

information on which use of the experimental method relies should be even approximately true?

But one can go even farther and flatly deny that the experimental method ought to be considered as being even *partially* instrumentally reliable. Fine (1986, 119) argues that the 'problem for the realist is how to explain the *occasional success* of a strategy that *usually fails*': the alleged instrumental reliability of the experimental method that Boyd takes as his *explanandum* is claimed to be nothing more than an artifact of 'Whiggish' history of science. Even if the use of the experimental method does lead, every now and then, to some notable successes,

> Overwhelmingly, the results of the conscientious pursuit of scientific inquiry are failures: failed theories, failed hypotheses, failed conjectures, inaccurate measurements, incorrect estimations of parameters, fallacious causal inferences, and so forth.
>
> (Fine 1986, 119)

Even though this picture of the average results of scientific inquiry could be suspected of downplaying the actual achievements of the experimental method, Fine surely puts his finger on a distinctive difficulty for Boyd's 'meta-methodological' strategy: by presupposing the instrumental reliability of the experimental method, Boyd's argument is hostage to the vicissitudes of a much more onerous assumption than usual 'no-miracle' arguments based on the instrumental success of carefully selected theories in 'mature' science. Boyd cannot simply assume that the instrumental reliability of the experimental method will be granted by positivistically-minded philosophers: he needs to argue that it *ought* to be.

There is however a third objection to Boyd's argument for scientific realism that seems to me even more disruptive. This objection, which is due to Laudan (1981, 45-46) and Fine (1986, 114-115), focuses on the form, rather than the premises, of the argument. The problem with Boyd's vindication of scientific realism is that it has just the same form as the sort of abductive inference which it is supposed to justify. What the anti-realist is in the business of denying is that one is entitled to infer that scientific theories are (approximately) true from the fact that they are predictively successful. But now, the sort of argument by which Boyd tries to vindicate this realist claim has precisely the form of that inference to the best explanation that the anti-realist rejects as unjustified. Boyd's attempt to establish the truth of scientific realism by claiming that it

represents the best explanation of the instrumental reliability of the scientific method is therefore just as good (or as bad) as the ground-level attempt to establish the truth of a scientific theory by claiming that the theory's being true is the best explanation of its being predictively successful. If the anti-realist rejects the argument at the ground-level (for example by appealing to Laudan's pessimistic meta-induction), it is not clear why she should accept it at any other level.

Moreover, Boyd's abduction from the instrumental reliability of the experimental method to the truth of scientific realism has hardly any *independent* evidence in its favour. The explanation of the instrumental reliability of the experimental method afforded by scientific realism is, unlike good scientific explanations, incapable of being independently tested. This seems to me to have bad consequences indeed for Boyd's claim that scientific realism ought to be seen as a very general *scientific* hypothesis which can be justified by a single application of that very method by which standard scientific hypotheses are justified.

My conclusion is then that, despite the appeal it can wield on naturalistically-minded philosophers, Boyd's argument for scientific realism fails to vindicate the claim that our present picture of the world cannot be too far from the truth. (Perhaps I should qualify my rejection of Boyd's argument by saying that it should not be taken to imply positivistic sympathies on my part; for one thing, I share his belief that empirically equivalent theories are not necessarily evidentially indistinguishable; but I think that his argument for the truth of scientific realism is hopelessly flawed). The availability of a proof that our present picture of the world cannot be too far from the truth is not, however, a precondition for the success of my attempt to show that commitment to correspondence truth need not make knowledge and cognitive progress impossible. As I have tried to argue, one can do very well without the epistemological reassurance promised by the brands of evolutionary epistemology and scientific realism discussed in this and the previous section.

6.7 On the Aim of the Game

It is sometimes argued, e.g. in Putnam (1982) and Ellis (1990), that the kind of naturalism involved in the approach to cognitive prescriptions described in this chapter

is refuted by the impossibility of formalizing 'reason' and/or by the intrinsic 'epistemic value' of truth.

Putnam (1982) argues that there can be no general algorithm for scientific inquiry because the kind of 'norms' (reasons) upon which human beings in general, and scientists in particular, base their cognitive decisions exhibit endless variety and plasticity[3]. There is no single canon of methodology to be slavishly followed:

> For every culture has norms which are vague, norms which are unreasonable, norms which dictate inconsistent beliefs. [...] Our task is not to mechanically *apply* cultural norms, as if they were a computer program and we were the computer, but to interpret them, to criticize them, to bring them and the ideals which inform them into reflective equilibrium.
>
> (Putnam 1982, 14)

The crucial point is that, while Putnam acknowledges the necessity of criticizing the (cognitive) norms dictated by one's present culture, he claims that the process of their criticism cannot itself be regimented into an algorithm to be mechanically applied. In this sense, reason 'transcends' all (cognitive) activities and institutions, in so far as it represents a regulative idea that we use in criticizing all sorts of norms, including second-order norms for the criticism of the (cognitive) activities and institutions of our present culture (Putnam 1982, 8). It is this 'transcendence' of reason that Putnam argues precludes a naturalistic approach to the vindication of methodological rules. The evaluation of our cognitive methods cannot be a matter of assessing their effectiveness for the pursuit of our cognitive aims. For if this were how our cognitive methods ought to be evaluated, effective cognitive rules would have to provide algorithms, or at least recipes, for cognitive success, and reason could not display its characteristic transcendence and open-endedness.

As it stands, this argument is incapable of refuting the sort of naturalistic approach advocated in this chapter. For I am not in the business of constructing a Grand Normative Theory. The claim that we actually possess such a Theory would certainly conflict with the fact that the sort of 'reasons' we commonly provide in the discussion of our beliefs and theories display a characteristic variety and open-endedness. However, I have merely attempted to show that the effectiveness of local and content-specific

---

[3] A discussion of the disappointing outcome of various attempts to regiment the practices of (successful) science into a universally binding scientific 'method' can be found in Putnam (1981, 188-200).

methodological rules can in many cases be vindicated on a posteriori grounds, and that such vindication need not consist in a proof of their effectiveness in all possible worlds to be instrumental in the production of significant improvements in our present epistemic situation. Therefore, the impossibility of a complete 'formalization' of reason is in itself no refutation of the approach defended in this chapter.

However, there is more to Putnam's argument than the realization of the transcendence and open-endedness of 'reason'. For in his theory this is not left as a brute fact, but is accounted for by the claim that truth cannot be described as correspondence to a knowledge-independent reality. According to Putnam, our only 'vital and working' notion of truth is in terms of the notion of rational acceptability, which notion is essentially evaluative. When we say that a proposition is true, we are not describing some non-evaluative fact about that proposition (Putnam believes that there are no facts independent of values). Rather, we are claiming that it would be rationally acceptable under sufficiently good epistemic conditions; in short, we are evaluating it. But then, if our only notion of truth is in terms of the notion of rational acceptability, it will be clearly impossible to use the former notion to provide a non-empty definition of the latter, and the notion of 'reason' will have to be taken as an inherently evaluative primitive.

Brian Ellis (whose rejection of the fact/value dichotomy is less explicit than Putnam's own) takes a similar line of argument when he contends that 'truth cannot be identified with any property or relationship which lacks intrinsic epistemic value' (Ellis 1990, 187). The reason is supposed to be that if truth were a property or relationship which held independently of our epistemic values, there would be no explanation 'of why rationally we should seek to discover what is true, rather than, say, round or married' (Ellis 1990, 188). To identify truth with a property of statements or beliefs is, he maintains, to commit the epistemic equivalent of the naturalistic fallacy: 'it is to treat an expression of epistemic approval as though it were the attribution of a property to the object of approval' (Ellis 1990, 187). Ellis' rationale for describing truth as inherently evaluative is not that since the correspondence-relation is, in some sense, incoherent, truth will have to be defined in epistemic — hence, evaluative — terms. His rationale for describing truth as inherently evaluative is quite simply that it is *necessarily* what we should rationally seek to discover. Accordingly, if the notion of truth cannot be seen as referring to the property of those statements that 'correspond' (in the sense defined in

section 1.4) to reality, the instrumental notion of 'reason' involved in the attempt to vindicate methodological rules on a posteriori grounds will not be definable in its terms.

I am not convinced by Ellis' rationale for seeing truth as inherently evaluative. Apart from well-known questions about the actual fallaciousness of the 'naturalistic fallacy', it is far from clear that truth is, in Ellis' phrase, *what we should rationally seek to discover*. Truth is surely *part* of what we should rationally seek to discover (this is why I describe *cognitive* aims as inherently *truth-linked* aims), but the goal of our cognitive efforts cannot be mere truth-acquisition and error-avoidance, if only because these goals will often pull in opposite directions. Moreover, there are uninteresting truths we have no reason at all to seek to discover (e.g., that the number of molecules in a mole of any substance — Avogadro's number — is at least twice as large as the number of members of the United States of America). It is not truth as such that is valuable, but those answers, explanations, and predictions (or what have you: fill in the details of your favoured picture of the nature of our cognitive aims) the possession of which we have some reason — whether pragmatic or theoretic — to value. Truth is more properly seen as an adequacy requirement on answers, explanations and predictions than as an independent goal of our inquiries. But then it can very well be seen as an objective property of statements or beliefs. The reason why we should seek to discover true answers, explanations and predictions is quite simply that just as only round balls are *good* footballs (and this does not entail that 'roundness' cannot be an objective property of balls), only true answers, explanations and predictions are *good* answers, explanations and predictions (and this does not entail that 'truth' cannot be an objective property of statements or beliefs).

One can of course ask *why* answers, explanations and predictions must be (correspondence-)true rather than, say, round or married to be good. Various pragmatic reasons can be produced. True predictions are preferable to false ones because they enable us to devise and adopt effective means to our ends. True answers and explanations are pragmatically preferable to false ones (i) because of the role they play in the formulation of successful predictions; (ii) because, assuming our cognitive efforts to be reasonably successful, they are less likely to be overturned by new evidence. A purely pragmatic account of the desirability of *true* answers, explanations and predictions is of course open to the criticisms (i) that suitable substitutes (e.g., empirical adequacy) could perform the above-mentioned pragmatic functions at least as well as truth and (ii)

that knowledge can be sought for its own sake. But even if the ultimate reason why answers, explanations and predictions must be true to be good should be, 'Because human beings *believe* they must be', or, 'Because human beings *prefer* true answers, explanations and prediction to false ones', this would not suffice to prove that truth cannot be an objective property of statements or beliefs: does the fact that kids prefer sweet foods to bitter ones prove that the disposition to taste sweet to members of the human species cannot be an objective property of things?

Since Ellis' argument is defective and Putnam's only reason for his commitment to an evaluative notion of truth lies in the claim that correspondence truth (which he takes to be the only serious alternative to epistemic truth) is incoherent[4], my conclusion is that no positive argument for seeing truth as inherently evaluative has been provided and that the naturalistic approach to the vindication of methodological rules is not undermined by arguments of the kind described in this section.

But one can, I think, take a further step and question the plausibility of the picture of the status of cognitive prescriptions involved in the acceptance of epistemic truth. For commitment to an inherently evaluative view of truth reverses the relation one is intuitively inclined to posit between truth and method, or, in other terms, between the theory of truth and epistemology. Ellis is very clear on this point:

> Every theory of truth which identifies truth with what it is right to believe depends on some epistemology. For it depends on one's theory of right belief. Thus an empiricist epistemology will yield an empiricist theory of truth; a rationalist theory of knowledge will yield a rationalist theory; and a coherentist epistemology will yield a coherence theory of truth.
>
> (Ellis 1990, 219)

Putnam holds similar views, even though he would probably qualify the statement that 'every theory of truth which identifies truth with what it is right to believe depends [...] on one's theory of right belief' by emphasizing that it is not one's *concept* of truth that depends on one's theory of right belief, but only one's *conception* of truth (i.e., the content of one's beliefs about it).

---

[4] Putnam rejects the fact/value dichotomy on the ground that truth cannot be an objective property of beliefs or statements, and not *vice-versa*. For a brief formulation of the argument, see Putnam (1990), 115-117; for a longer formulation, see Putnam (1981), ch. 6.

But is it really plausible to take such epistemic notions as 'justified' or 'rationally acceptable' as primitives and build one's characterization of truth upon them? Alvin Goldman (1986, 147 f.) contends that epistemic notions are far more in need of explication than truth and that, given the *prima facie* plausibility of the traditional view that 'justification' and 'rational acceptability' are to be explained with reference to truth, 'the onus is on Putnam, Dummett, or like-minded theorists to show that these notions have non-truth-linked explications'. The reason why I share Goldman's contention is, essentially, that one can make better sense of the fact that people make judgements on the cognitive value of competing methodological rules by endorsing a non-epistemic theory of truth than an epistemic one.

If the aim of our cognitive efforts is to discover *true* answers, explanations and predictions, then of course it will make sense to ask which methodological rules are cognitively most effective: what we want to know is simply which methodological rules are more likely to lead us to the acceptance of true answers, explanations and predictions. But this account is not available to supporters of epistemic truth. For there would be no point in assessing competing sets of methodological rules with respect to their truth-conduciveness if truth were understood in an epistemic fashion, i.e., if it were defined either in terms of a given set of methodological rules (Ellis) or in terms of a primitive, non-formalizable concept of 'reason' (Putnam). If truth simply *is* what our methodological rules or our 'reason' would lead us to accept in epistemically ideal conditions, then it will be obviously circular to use our concept of truth in assessing the cognitive value of our methodological rules or the 'rationality' of our reason. The very possibility of appraising the cognitive value of competing sets of methodological rules or styles of reasoning appears to require that truth be understood independently of any given set of methodological rules or style of reasoning (an alternative could be to replace truth with 'pragmatic utility' or other such things as a criterion for the appraisal of competing methodologies or styles of reasoning; but this is neither Putnam's nor Ellis' strategy, because it is evident from their writings that they are not trying to 'reduce' truth to such non-evaluative notions).

To be sure, upholders of epistemic truth can make sense of *local* processes of methodological evaluation and/or develop viable accounts of the reliability of *specific* cognitive methods. This is because local processes of methodological evaluation and accounts of the reliability of specific cognitive methods need not always affect the

fundamental epistemic values and/or methods involved in the characterization of (epistemic) truth. Putnam (1981, 132), for example, argues that we can account for the reliability of our perceptual knowledge by describing «how our perceptions result from the operation of transducing organs upon the external world». He is entitled to do so because such an account involves a sufficiently circumscribed modification of our total system of beliefs, a consistent part of which — most notably, the belief that our perceptual knowledge *is* reliable — is taken for granted in accounting for the reliability of our perceptual apparatus.

However, when it comes to the comparative evaluation of more comprehensive sets of methodological rules (or styles of reasoning), one cannot always rely on the availability of such an uncontroversial body of belief to arrive at a verdict. It is not difficult to imagine mutually incompatible but internally consistent sets of methodological rules (or styles of reasoning) the choice between which upholders of epistemic truth are bound to regard, ultimately, as a matter of (cognitive) taste. Think, for example, of the style of reasoning of present-day science, as opposed to the ludicrously conservative style of reasoning described in section 1.1 above. Upholders of correspondence truth can claim — though of course they cannot prove to the satisfaction of the sceptic — that the former is a much more reliable means to cognitive success then the latter (that the latter may be a more reliable means for the fixation of belief is irrelevant to our issue). Upholders of epistemic truth, on the contrary, cannot justify a preference for the former style of reasoning by appealing to its truth-conduciveness, because they have no independent notion of truth in terms of which the effectiveness of such mutually incompatible but internally consistent sets of methodological rules (or styles of reasoning) could be compared. Granted, upholders of correspondence-truth cannot give a *proof* of the effectiveness of the style of reasoning of present-day science any more than upholders of epistemic truth can. But the point is that upholders of epistemic truth do not have even the *concepts* to formulate the right question, which has to do with the comparative truth-conduciveness of competing styles of reasoning rather than with the choice between different conceptions of truth. It is small wonder, then, that both Putnam and Ellis end up giving extravagant answers to the issue of what would provide a justification of our most fundamental cognitive values and/or methods.

Ellis' view of epistemology is that of a 'values-based' theory 'on which our inductive practices either turn out to be rational according to the theory, or not rational,

but for reasons we are persuaded are correct'. This theory consists, in short, in a model of the cognitive behaviour of an 'ideally rational being'. But to the question, Why should we accept the cognitive values of such a being?, Ellis' (1990, 261) only answer is: '"Because you are human" [...] these *are* your epistemic values, and you cannot, like a god, step outside your value system to judge whether or not it is rational to have them'. In a similar vein, Putnam's (1981, 134) vindication of the ideal of rational acceptability characteristic of Western science is, ultimately, that that ideal 'is *part of our idea of human cognitive flourishing*, and hence part of our idea of total human flourishing, of Eudaemonia': take it or leave it.

Even if I have no proof that the only way to settle global disagreements about our most fundamental cognitive values and/or methods is by subscribing to a correspondence theory of truth, the alternatives advocated by Ellis and Putnam seem to me very unsatisfactory. As I have already remarked, our characterization of the aims of rational inquiry cannot but be the upshot of a trade-off between competing cognitive values. And it is a fact that there cannot be, in Quine's phrase, any 'point of cosmic exile' from which to adjudicate the effectiveness of competing sets of methodological rules or styles of reasoning. But when compared with the accounts available to upholders of epistemic theories of truth, the possibility of seeing our cognitive efforts as intentionally directed to the discovery of (correspondence-)true answers, explanations and predictions surely provides a better explanation of our intuition that the style of reasoning of present-day science is our best means to the acquisition of factual knowledge of the world.

Realism Without a God's Eye View

I have attempted to show that acceptance of the correspondence theory of truth need not make truth epistemically inaccessible and knowledge impossible by arguing that an accurate blend of (naturalistic) externalism in the theory of knowledge and (pragmatic) internalism in the theory of epistemic justification can provide a satisfactory account of our epistemic access to knowledge-independent reality without committing us to assume that human beings have access to a God's eye view of that reality.

The kind of epistemic access to reality that it is possible to ascribe to human beings on the basis of the approach advocated in this work involves no certainty that the states of affairs that one can know to obtain are indeed as one believes them to be. In other terms, an important feature of our epistemic access to reality has been argued to be its *fallibility*. On the other hand, the epistemic access to reality I have claimed human beings can be credited with involves much more than mere 'thin' epistemic justification. S's knowing that *p* involves an appropriate factual connection of the knower with the world, such that S's belief that *p* can be taken to be a reliable indicator of the actual state of the world. It is precisely this 'naturalization' of knowing that enables metaphysical realists to say that no God's eye view of reality is required for us to have epistemic access to the way the world is 'in itself'. A modest degree of initial 'epistemic luck' can be sufficient to set in motion the self-corrective process that I have tried to describe in chapter 6. Of course we have no *proof* that our cognitive and methodological efforts constitute in fact such a virtuous process. (But neither do upholders of those internalist views of knowing which I have argued to be either a byproduct of old-

fashioned dogmatic epistemologies or the consequence of the mistaken belief that an internalist view of epistemic justification involves commitment to an internalist view of knowing). However, the fact that we cannot prove that our cognitive and methodological efforts constitute an epistemically successful process need not prevent us from learning many things about the world and about the reliability (or lack thereof) of our strategies for improving our epistemic access to it.

This may not seem an impressive result, but I have also argued that lowering our cognitive ambitions by substituting a merely 'phenomenal' world for the knowledge-independent reality we commonly take our beliefs to be about will not give us a more appealing view of our cognitive achievements. Admittedly, I have not even attempted to clear the correspondence theory of truth from the charges of incoherence that are recurringly levelled against it. However, I hope I have made it plausible that, if the conjunction of metaphysical realism and correspondence truth makes any sense at all, it need not create greater difficulties to the belief that we have some degree of epistemic access to reality than the conjunction of metaphysical anti-realism and epistemic truth.

A Model-Theoretic Approach to Comparative Verisimilitude

In section 6.7 I pointed out that, although truth plays an essential role in the definition of our cognitive aims, truth-seeking and error-avoidance cannot be the whole story about the goals of our cognitive efforts. Our cognitive efforts are directed to the acquisition of *informative* truths; and for the sake of acquiring such truths we may be willing to follow cognitive strategies involving a considerable risk of error. Indeed, there can be no doubt that false theories such as Newtonian Mechanics fulfil our cognitive aims much better than tautologies such as 'All bodies are either extended or unextended'. This is why some writers have suggested that the truth-related component of our cognitive aims may have more to do with the maximization of the 'verisimilitude' (closeness to truth) of our theories, than with mere truth-acquisition and error-avoidance.

The relevance of the notion of verisimilitude, or truthlikeness, for a realist conception of knowledge was first noticed by Karl Popper when he realized that the legitimacy of scientific judgements about the comparative verisimilitude of competing theories was of primary importance for his falsificationist epistemology. For a central claim of Popper's falsificationism is that even the best corroborated theories of physical science are most likely false. Accordingly, only the idea that a false theory can legitimately be described, in certain circumstances, as more 'truthlike' than another seems to be able to reconcile Popper's view of the history of science as a sequence of bold conjectures and decisive refutations with his faith in the rationality and cognitively progressive character of scientific inquiry.

But of course the relevance of the notion of verisimilitude is not circumscribed to Popperian falsificationism. Indeed, Oddie (1986, ix) argues that the issue of verisimilitude, far from being an exclusive concern of popperian falsificationists, plays an essential role in any epistemology committed to the modest 'realist' assumptions that the aim of an inquiry, as an inquiry, is the truth of some matter, and that one false theory may realise this aim better than another. For it is a consequence of these two assumptions that it must be perfectly meaningful to say that a false theory is closer to the truth than another theory, which in turn raises the problem of developing an analysis of the precise meaning of judgements of comparative verisimilitude among theories.

Popper's original treatment of the issue of verisimiltude included a qualitative definition of comparative verisimilitude and a probabilistic measure of absolute verisimilitude (see Popper 1963, ch. 10 and Addenda; 1972, chs. 2 and 9). Popper would describe his definitions as 'semantic' because he saw them as a development of Tarski's 'semantic' conception of truth. However, since Popper's definitions presuppose the classical view of theories as sets of sentences, and moreover their key-notions — those of the truth content and the falsity content of a theory — are cast in terms of Tarski's notion of a 'consequence class', Popper's proposal is more commonly described as 'syntactic'. Both its parts have unfortunately been shown to be inadequate. For Popper's definition of comparative verisimilitude has the devastating consequence that no false theory can be closer to the truth than any other theory (Miller 1974, 170-174; Tichý 1974, 156 f.), and his measure of absolute verisimilitude is such that it makes possible to settle which one of two false theories is closer to the truth without any factual knowledge over and above the knowledge that the two theories are indeed false (Tichý 1974, 158).

A different approach to verisimilitude was developed by Pavel Tichý and Graham Oddie at Otago, and independently by Risto Hilpinen and Ilkka Niiniluoto in Finland. This is the so-called 'similarity' approach, which purports to give a solution to the problem of verisimilitude by developing a definition of distance between theories. The starting-point of the definitions of verisimilitude developed in this tradition is, typically, the notion of similarity between constituents of first order languages. This approach was first criticized by Miller (1974, 175-177) for its failure to characterize a translation-invariant concept of verisimilitude and it is now highly doubtful if the 'similarity' approach can provide anything resembling the 'objective' notion of verisimilitude

Popper's original definitions were meant to capture (that is, unless it is complemented by further considerations about 'right' and 'wrong' sets of traits for describing the world). This of course does not suffice to show that the 'similarity' approach is fundamentally flawed (one can either maintain that it is merely incomplete, or argue that Miller's translation-invariance requirement is unjustifiably strong), but it does provide a motivation for investigating alternative and possibly more objective frameworks for the purpose of defining verisimilitude.

The framework investigated in this Appendix is provided by the so-called 'semantic' view of theories (important work on the notion of verisimilitude in the framework of the semantic view of theories has been done by Theo A. Kuipers, but my approach will be substantially different). I propose to use the machinery of the semantic view of theories to provide necessary and sufficient conditions for the truth of sentences of the form, 'B is truth-increasing with respect to A', where A and B are theories that can be expressed in first order language. By doing this, I hope to provide a model-theoretic definition of something very close to the comparative (as distinct from the classificatory and the quantitative) notion of verisimilitude (the relation thus defined is meant to satisfy the usual ordering postulates, that is to say, it is meant to be irreflexive and transitive). I use the phrase, 'B is truth-increasing with respect to A' rather than, say, 'B approximates the truth better than A', because I want to emphasize that my definition does not take into account issues of *accuracy*, but only issues of *comprehensiveness*. In this respect my proposal is wholly in the spirit of Popper's original approach. The reason why, unlike Popper, I do not say that my definition purports to capture the comparative notion of verisimilitude, but only something (hopefully) very close to it, is precisely that I believe that a fully satisfactory account of the truth-related component of scientific progress would have to address issues of accuracy as well as of comprehensiveness.

1. *Preliminaries*

According to standard first order logic semantics, an interpretation I of a formalized language L is given, roughly, by a $n+1$-tuple S consisting of a set of objects (the domain of the interpretation, D) and of a set of relations among those objects:

$$S = <D,R_1,\ldots,R_n>$$

I assigns set D to each individual variable of L as its range, and relations $R_1,\ldots,R_n$ to the predicate letters $P_1,\ldots,P_n$ of L as their denotations. (We can ignore the complications connected with the interpretation of names and function symbols).

Now, an interpretation is an interpretation of a language, but the set-theoretical entity *S* which determines one of the possible interpretations of L can be considered *per se*, independently of any particular language, and I shall call such an autonomous ordered tuple a *structure*. Structures are important in the analysis of scientific theories because one need not conceive a scientific theory as being in the first place an axiomatic system which is subsequently given an appropriate interpretation. Apart from the disappointment of neo-positivistic hopes of using formalization for distinguishing the genuinely factual from the merely conventional components of scientific theories, formalization can hardly provide us with a better understanding of the nature of scientific theories, for the simple reason that for most scientific theories there will be no axiomatic system capable of picking out just those structures the theory was originally purported to describe (if the theory contains the real number continuum, this is an immediate consequence of the Löwenheim-Skolem theorems). Therefore, we may wish to focus our attention on the structures the theory purports to pick out without any further syntactic mediation. This is what the so-called *semantic* view of theories (see, e.g., Suppes 1960; Przelecki 1969; Van Fraassen 1980 and 1985; Giere 1988) is meant to do, and this is, as I have said, the approach that I propose to adopt in this paper.

A scientific *theory* can thus be described as a set of structures. The following definitions help to articulate this semantic view of theories.

Two structures $S = <D,R_1,\ldots,R_n>$ and $S' = <D',R'_1,\ldots,R'_n>$ are *isomorphic* if there exists a one-one function f which maps D on D' in such a way that

$$\forall x_1,\ldots,x_k \epsilon D[R_i(x_1,\ldots,x_k) \leftrightarrow R'_i(f(x_1),\ldots,f(x_k))]$$

A structure $S' = <D',R'_1,\ldots,R'_n>$ is an *extension* of a structure $S = <D,R_1,\ldots,R_n>$ iff D is a subset of D' and

$$\forall x_1,\ldots,x_k \epsilon D[R'_i(x_1,\ldots,x_k) \leftrightarrow R_i(x_1,\ldots,x_k)]$$

A structure $S' = <D',R'_1,\ldots,R'_n>$ is a *proper* extension of a structure $S = <D,R_1,\ldots,R_n>$ iff D is a *proper* subset of D' and

$$\forall x_1,\ldots,x_k \in D[R'_i(x_1,\ldots,x_k) \leftrightarrow R_i(x_1,\ldots,x_k)]$$

A structure $S' = <D',R'_1,\ldots,R'_k, S_1,\ldots,S_m>$ is an *expansion* of a structure $S = <D,R_1,\ldots,R_k>$ iff $D = D'$ and $R_i = R'_i$ for any i between 1 and k. A structure $S' = <D',R'_1,\ldots,R'_k, S_1,\ldots,S_m>$ is a *proper* expansion of a structure $S = <D,R_1,\ldots,R_k>$ iff $D = D'$, $m > 0$, and $R_i = R'_i$ for any i between 1 and k.

A structure *S* is a *substructure* of a structure *S'* iff *S'* is either an extension, or an expansion, or an extension of an expansion, or an expansion of an extension of *S*. A structure *S* is a *proper* substructure of a structure *S'* iff *S'* is either a proper extension, or a proper expansion, or an extension of a proper expansion, or an expansion of a proper extension of *S*.

I shall say that a structure *S* can be (properly) *embedded* in a structure *S'*, or that a structure *S'* (properly) *embeds* a structure *S*, iff *S* is a (proper) substructure of *S'*.

If we are willing to say that the world is made up of objects, (properties) and relations, we can also say that the world is a set of objects standing in certain mutual relations. We can describe the world as the set of those objects which belong to the domain D, and stand in the relations $R_1,\ldots,R_n$, of a certain structure $S = <D,R_1,\ldots,R_n>$. We can say that the world *is* a structure. Or we can make the weaker claim that there is just one structure that corresponds to the (actual) world, or, equivalently, that the (actual) world *defines* just one structure[1]. For the sake of brevity, I will henceforth talk of the actual world as if it were a structure. But a reader who doesn't like this idea can substitute in my tentative definition of comparative verisimilitude the phrase '*X* can be embedded in the structure defined by the actual world' for the phrase '*X* can be embedded in the actual world'.

A structure *A'* is a (proper) *substructure* of a theory A iff it is a (proper) substructure of some structure of A. Where *A* and *B* are structures and *A* can be embedded in *B*, the *content-excess* of *B* over *A* is defined as the set of those structures

[1] This claim clearly presupposes a realist view of objects, (properties) and relations. If these should be taken to be language-dependent, then there would be no one structure defined by the (actual) world, but many different and incommensurable structures defined by many different and incommensurable worlds.

(other than *B*) in which *A* can be embedded that are neither embeddable in, nor can embed *B* (the content-excess of *B* over *A* is, so to speak, the set of those 'possible worlds' which are compatible with *A* but not with *B*).

A theory B *constrains* the content-excess of *B'* over *A'* iff some structure in the content-excess of *B'* over *A'* is neither a substructure of B, nor embeds any structure of B (a theory B that constrains the content-excess of *B'* over *A'* 'prohibits', so to speak, some of the possible worlds which are compatible with *A'* but not with *B'*).

A substructure *B'* of a theory B is an *informative embedding* of a substructure *A'* of a theory A iff (i) *A'* can be properly embedded in *B'*; (ii) *B'* can be embedded in the actual world; (iii) B constrains the content-excess of *B'* over *A'*.

2. *The Definition*

Where A and B are theories, *B is truth-increasing with respect to A*

iff

a) for any substructure *A'* of A which can be embedded in the actual world, there is some substructure *B'* of B, such that *B'* is an informative embedding of *A'*

and

b) there is some substructure *B'* of B, such that *B'* can be embedded in the actual world and there is no substructure *A'* of A, such that *A'* is an informative embedding of *B'*.

The notion of 'truth-increase' introduced by this definition is such that if there is some truth-loss in the passage from A to B, B cannot be closer to the truth than A.

The Tichý-Miller argument against Popper's definition of comparative verisimilitude cannot be adapted to undermine the present proposal, which does not require that the falsity content of a theory B which is closer to the truth than a theory A be a (proper or improper) subset of the falsity content of A. That a similar requirement should not be included in a definition of comparative verisimilitude seems to descend from the fact that better theories get closer to the truth by making more numerous and more precise claims about reality. If theory B has something to say on things theory A is non-committal about, no wonder that B will be likely to make a

number of wrong claims which are not made by A. But that does not entail that B cannot be closer to the truth than A!

Miller's argument against the 'similarity' approach to verisimilitude also fails to affect the present proposal. This is because the present proposal presupposes a realist view of objects, (properties,) and relations. If one is willing to subscribe to this kind of realism, the definition will ensure the translation-invariance of the judgements of comparative verisimilitude it entails.

The present definition encompasses any kind of truth-increase, including trivial, or *horizontal*, truth-increase (e.g., *General Relativity* + *the snow is white* comes out as truth-increasing with respect to *General Relativity* alone). However, if we add the further requirement that there be some substructure of (some conservative logical weakening of) B which can be embedded in the actual world in which no structure of A can be embedded, we get something as a definition of non-trivial, or *vertical*, truth-increase. One has a non-trivial, or *vertical* truth-increase when B corrects A while preserving its 'truth content' (hence no theory can be *vertically* truth-increasing with respect to a true theory).

If theory B is closer (even *vertically* closer) to the truth than theory A, one can devise a theory A' such that B is neither vertically nor horizontally closer to the truth than A' by adding to A an arbitrary true claim p which is not made by B: suppose that *General Relativity* (B) is truth-increasing with respect to *Newtonian Mechanics* (A); then *General Relativity* will fail to be truth-increasing with respect to *Newtonian Mechanics* + *the snow is white* (A'). This of course does not imply that A' will be truth-increasing, let alone *vertically* truth-increasing, with respect to B. Moreover, if p is consistent with B, it will be child's play to devise a B' which does get closer to the truth than A': simply add p to B as well. If, on the contrary, p is inconsistent with B, it is clear that A' makes some true claim about the world which a consistent extension of B cannot possibly make. If that is the case, it seems reasonable to say that B is *not* closer to the truth than A', since there is some truth-loss in the passage from A' to B.

Starting from a theory A, one cannot get a theory B which is truth-increasing with respect to A merely by adding a false claim f to A. By taking into account only those substructures of the conservative logical weakenings of our theories which can be embedded in the actual world we do get rid of the false consequences of our theories while retaining their true ones, but we are prevented from taking illegitimate advantage

from this charitable attitude by the requirement that the relevant conservative logical weakening of theory B *constrain* the content-excess of *B'* over *A'*. This requirement precludes a theory wrongly claiming that f from coming out as truth-increasing with respect to a theory which is silent on whether f or not-f merely because of its greater logical strength. Accordingly, it does not follow from the present definition that, if A and B are false theories, B is truth-increasing with respect to A if and only if A is a logical consequence of B and not vice versa.

3. *Conclusion*

It is an obvious consequence of my analysis that the formulation of theories which are truth-increasing (even vertically truth-increasing) with respect to their predecessors cannot be the only aim of science. Cognitive virtues such as accuracy (which is part of the truth-related component of our cognitive aims) and explanatory power (which is perhaps more pragmatically-oriented) also have to be taken into consideration in the characterization of scientific progress, which means that mere truth-increase cannot be seen as providing either a sufficient or a necessary condition for cognitive progress. However, the purpose of this Appendix was not to present a full characterization of the aims of science, but only to show how the machinery of the semantic conception of theories can be used to develop a consistent view of what it could mean to say that a (false) theory is closer to the truth than another theory.

References

Alston, W.A. 1989. *Epistemic Justification. Essays in the Theory of Knowledge*, Ithaca-London: Cornell University Press.

Armstrong, D.M. 1973. *Belief, Truth and Knowledge*, London: Cambridge University Press.

Austin, J.L. 1946. 'Other Minds', *Proceedings of the Aristotelian Society*, Supplementary vol. 20, 148-187.

Austin, J.L. 1962. *Sense and Sensibilia*, Oxford: Oxford University Press.

Blackburn, S. 1984. *Spreading the Word*, Oxford: Clarendon Press.

Bloor, D. 1976. *Knowledge and Social Imagery*, London: Routledge and Kegan Paul.

Bonjour, L. 1980. 'Externalist Theories of Empirical Knowledge', *Midwest Studies in Philosophy* 5, 53-73.

Bonjour, L. 1985. *The Structure of Empirical Knowledge*, Cambridge Mass.-London: Harvard University Press.

Boyd, R. 1973. 'Realism, Underdetermination and a Causal Theory of Evidence', *Nous* 7, 1-12.

Boyd, R. 1983. 'On the Current Status of the Issue of Scientific Realism', *Erkenntnis* 19, 45-90.

Boyd, R. 1985. 'Lex Orandi est Lex Credendi', in P.M. Churchland and C.A. Hooker eds. *Images of Science*, Chicago-London: The University of Chicago Press, 3-34.

Burge, T. 1979. 'Individualism and the Mental', in P. French, T. Ühling and H. Wettstein eds. *Midwest Studies in Philosophy 4: Studies in Metaphysics*, Minneapolis: University of Minnesota Press, 73-122.

Campbell, D.T. 1974. 'Evolutionary Epistemology', in P.A. Schilpp ed. *The Philosophy of Karl Popper*, 2 vols., La Salle Ill.: Open Court, 413-463.

Campbell, D.T. 1986. 'Science Policy from a Naturalistic Sociological Epistemology', *PSA 1986, vol. 2*, ed. by P.D. Asquith and P. Kitcher, East Lansing Mich.: Philosophy of Science Association, 14-29.

Cartwright, N. 1994. 'Fundamentalism vs. the Patchwork of Laws', *Proceedings of the Aristotelian Society* n.s. 94, 279-292.

Chisholm, R. 1989. *Theory of Knowledge*, 3rd edition, Englewood Cliffs N.J.: Prentice-Hall.

Cohen, S. 1991. 'Skepticism, Relevance, and Relativity', in B.P. McLaughlin ed. *Dretske and his Critics*, Oxford: Blackwell, 17-37.

Craig, E. 1989. 'Nozick and the Sceptic: the Thumbnail Version', *Analysis* 49, 161-162.

Dancy, J. 1985. *An Introduction to Contemporary Epistemology*, Oxford: Blackwell.

Danto, A.C. 1967. 'On Knowing That We Know', in A. Stroll ed. *Epistemology: New Essays in the Theory of Knowledge*, New York: Harper & Row, 32-53.

Davidson, D. 1986. 'A Coherence Theory of Truth and Knowledge', in E. LePore ed. *Truth and Interpretation*, Oxford: Blackwell, 307-319.

Davidson, D. 1990. 'The Structure and Content of Truth', *The Journal of Philosophy* 87, 279-328.

Descartes, R. 1984-85. *The Philosophical Writings*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, 2 vols., Cambridge: Cambridge University Press.

Devitt, M. 1991a. 'Aberrations of the Realism Debate', *Philosophical Studies* 61, 43-63.

Devitt, M. 1991b. *Realism and Truth*, 2nd edition, Oxford: Blackwell.

Dretske, F. 1977. 'Laws of Nature', *Philosophy of Science* 44 (1977), 248-268.

Dretske, F. 1981a. *Knowledge and the Flow of Information*, Oxford: Blackwell.

Dretske, F. 1981b. 'The Pragmatic Dimension of Knowledge', *Philosophical Studies* 40, 363-378.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*, Cambridge Mass.-London: The MIT Press.

Dretske, F. 1991. 'Replies', in B.P. McLaughlin ed. *Dretske and his Critics*, Oxford: Blackwell, 180-221.

Dummett, M. 1976. 'What is a Theory of Meaning? (II)', in G. Evans and J. McDowell eds. *Truth and Meaning*, Oxford: Clarendon Press, 67-137.

Dummett, M. 1978. *Truth and Other Enigmas*, London: Duckworth.

Dummett, M. 1982. 'Realism', *Synthese* 52, 55-112.

Ellis, B. 1965. 'A Vindication of Scientific Inductive Practices', *American Philosophical Quarterly* 2, 296-304.

Ellis, B. 1988. 'Internal Realism', *Synthese* 76, 409-434.

Ellis, B. 1990. *Truth and Objectivity*, Oxford: Blackwell.

Feldman, R. 1981. 'Fallibilism and Knowing that One Knows', *The Philosophical Review* 90, 266-282.

Field, H. 1972. 'Tarski's Theory of Truth', *The Journal of Philosophy* 69, 347-375.

Field, H. 1982. 'Realism and Relativism', *The Journal of Philosophy* 79, 553-567.

Fine, A. 1986. *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago-London: The University of Chicago Press.

Foley, R. 1987. *The Theory of Epistemic Rationality*, Cambridge Mass.-London: Harvard University Press.

Frege, G. 1967. *The Basic Laws of Arithmetic*, trans. M. Furth, Los Angeles: University of California Press.

Friedman, M. 1985. 'Truth and Confirmation', in Kornblith 1985, 147-167.

Gettier, E. 1963. 'Is Justified True Belief Knowledge?', *Analysis* 23 (1963), 121-123.

Giere, R.N. 1988. *Explaining Science*, Chicago-London: The University of Chicago Press.

Goldman, Alan H. 1988. *Empirical Knowledge*, Berkeley-Los Angeles-London: University of California Press.

Goldman, Alvin I. 1976. 'Discrimination and Perceptual Knowledge', *The Journal of Philosophy* 73, 771-791.

Goldman, Alvin I. 1979. 'What is Justified Belief?', in G. Pappas ed. *Justification and Knowledge*, Dordrecht: Reidel, 1-23.

Goldman, Alvin I. 1980. 'The Internalist Conception of Justification', *Midwest Studies in Philosophy* 5, 27-51.

Goldman, Alvin I. 1986. *Epistemology and Cognition*, Cambridge Mass.-London: Harvard University Press.

Goodman, N. 1973. *Fact, Fiction, and Forecast*, 3rd edition, Indianapolis-New York: Bobbs-Merrill.

Goodman, N. 1978. *Ways of Worldmaking*, Hassocks: The Harvester Press.

Grayling, A.C. 1990. *An Introduction to Philosophical Logic*, 2nd edition, London: Duckworth.

Grayling, A.C. 1992. 'Epistemology and Realism', *Proceedings of the Aristotelian Society* n.s. 92, 47-65.

Haack, S. 1976. 'Is it True What They Say about Tarski?', *Philosophy* 51, 323-336.

Haack, S. 1978. *Philosophy of Logics*, Cambridge: Cambridge University Press.

Haack, S. 1979a. 'Fallibilism and Necessity', *Synthese* 41, 37-63.

Haack, S. 1979b. 'Epistemology *With* a Knowing Subject', *The Review of Metaphysics* 33, 309-335.

Haack, S. 1987. '"Realism"', *Synthese* 73, 275-299.

Hacking, I. 1983. *Representing and Intervening*, Cambridge: Cambridge University Press.

Hookway, C. 1990. *Scepticism*, London-New York: Routledge.

Horwich, P. 1982. 'Three Forms of Realism', *Synthese* 51, 181-201.

Horwich, P. 1990. *Truth*, Oxford: Blackwell.

Husserl, E. 1970. *Logical Investigations*, trans. J.N. Findlay, 2 vols., London-Henley: Routledge & Kegan Paul.

Jardine, N. 1986. *The Fortunes of Inquiry*, Oxford: Clarendon Press.

Kant, I. 1933. *Critique of Pure Reason*, trans. N. Kemp Smith, 2nd edition, London: Macmillan.

Kant, I. 1992. *Lectures on Logic*, trans. J.M. Young, Cambridge: Cambridge University Press.

Kim, K. 1993. 'Internalism and Externalism in Epistemology', *American Philosophical Quarterly* 30, 303-316.

Kinghan, M. 1986. 'The External World Sceptic Escapes Again', *Philosophia* 16, 161-66.

Kitcher, P. 1985. 'A Priori Knowledge', in Kornblith 1985, 129-145.

Kitcher, P. 1992. 'The Naturalists Return', *The Philosophical Review* 101, 53-114.

Kornblith, H. ed. 1985. *Naturalizing Epistemology*, Cambridge Mass.-London: The MIT Press.

Kripke, S. 1972. 'Naming and Necessity', in D. Davidson and G. Harman eds. *Semantics of Natural Language*, Dordrecht-Boston: Reidel, 253-355.

Kuipers T.A.F. 1982, 'Approaching Descriptive and Theoretical Truth', *Erkenntnis* 18,

343-378.

Lakatos, I. 1978. *The Methodology of Scientific Research Programmes. Philosophical Papers, vol. 1*, ed. by J. Worrall and G. Currie, Cambridge: Cambridge University Press.

Laudan, L. 1981. 'A Confutation of Convergent Realism', *Philosophy of Science* 48, 19-49.

Laudan, L. 1984. *Science and Values*, Berkeley-Los Angeles-London: University of California Press.

Laudan, L. 1989. 'If It Ain't Broke, Don't Fix It', *The British Journal for the Philosohy of Science* 40, 369-375.

Lehrer, K. 1990. *Theory of Knowledge*, London: Routledge.

Lipton, P. 1991. *Inference to the Best Explanation*, London-New York: Routledge.

Loar, B. 1987. 'Truth Beyond All Verification', in B.M. Taylor ed. *Michael Dummett: Contributions to Philosophy*, Dordrecht-Boston-Lancaster: Nijhoff.

Maffie, J. 1990. 'Recent Work on Naturalized Epistemology', *American Philosophical Quarterly* 27, 281-293.

McDowell, J. 1978. 'On "The Reality of the Past"', in C. Hookway and P. Pettit eds. *Action and Interpretation: Studies in the Philosophy of the Social Sciences*, Cambridge: Cambridge University Press, 127-144.

McGinn, C. 1984. 'The Concept of Knowledge', *Midwest Studies in Philosophy* 9, 529-554.

Miller, D. 1974. 'Popper's Qualitative Theory of Verisimilitude', *The British Journal for the Philosophy of Science* 25, 166-177.

Nagel, T. 1986. *The View From Nowhere*, New York-Oxford: Oxford University Press.

Newton-Smith, W.H. 1981. *The Rationality of Science*, London-New York: Routledge & Kegan Paul.

Niiniluoto, I. 1987. *Truthlikeness*, Dordrecht-Boston-Lancaster-Tokyo: Reidel.

Nozick, R. 1981. *Philosophical Explanations*, Oxford: Clarendon Press.

Oddie, G. 1986. *Likeness to Truth*, Dordrecht: Reidel.

Papineau, D. 1987. *Reality and Representation*, Oxford: Blackwell.

Peirce, C.S. 1934. *Collected Papers*, vol. 5, ed. by C. Hartshorne and P. Weiss, Cambridge: Harvard University Press.

Pettit, P. 1991. 'Realism and Response-Dependence', *Mind* 100, 587-626

Popper, K.R. 1959. *The Logic of Scientific Discovery*, London: Hutchinson.

Popper, K.R. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*, London: Routledge and Kegan Paul.

Popper, K.R. 1972. *Objective Knowledge*, Oxford: Clarendon Press.

Popper, K.R. 1990. *A World of Propensities*, Bristol: Thoemmes.

Prichard, H.A. 1950. *Knowledge and Perception*, Oxford: Clarendon Press.

Przelecki, M. 1969. *The Logic of Empirical Theories*, London: Routledge and Kegan Paul.

Putnam, H. 1975. *Mind, Language and Reality. Philosophical Papers, vol. 2*, Cambridge: Cambridge University Press.

Putnam, H. 1976. 'What is "Realism"?', *Proceedings of the Aristotelian Society* n.s. 76, 177-194.

Putnam, H. 1981. *Reason, Truth and History*, Cambridge: Cambridge University Press.

Putnam, H. 1982. 'Why Reason can't be Naturalized', *Synthese* 52, 3-23.

Putnam, H. 1983. *Realism and Reason. Philosophical Papers, vol. 3*, Cambridge: Cambridge University Press.

Putnam, H. 1988. *Representation and Reality*, Cambridge Mass.-London: The MIT Press.

Putnam, H. 1990. *Realism with a Human Face*, ed. by J. Conant, Cambridge Mass.-London: Harvard University Press.

Quine, W.V.O. 1953. *From a Logical Point of View*, Cambridge Mass: Harvard University Press.

Quine, W.V.O. 1969. *Ontological Relativity and Other Essays*, New York: Columbia University Press.

Radford, C. 1966. 'Knowledge — By Examples', *Analysis* 27, 1-11.

Reichenbach, H. 1949. 'The Logical Foundations of the Concept of Probability', in H. Feigl and W. Sellars eds. *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 305-323.

Rescher, N. 1973. *Conceptual Idealism*, Oxford: Blackwell.

Resnik, D.B. 1993. 'Critical Discussion of Ronald Giere, *Explaining Science: A Cognitive Approach*', *Erkenntnis* 38, 261-271.

Rorty, R. 1982. *Consequences of Pragmaticism (Essays: 1972-1980)*, Minneapolis: University of Minnesota Press.

Sellars, W. 1963. *Science, Perception and Reality*, London: Routledge and Kegan Paul.

Sosa, E. 1991. *Knowledge in Perspective*, Cambridge: Cambridge University Press.

Stich, S.P. 1985. 'Could Man Be an Irrational Animal? Some Notes on the Epistemology of Rationality', in Kornblith 1985, 249-267.

Strawson, P.F. 1952. *Introduction to Logical Theory*, London: Methuen.

Suppes, P. 1960. 'A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences', *Synthese* 12, 287-301.

Tarski, A. 1936. 'The Concept of Truth in Formalised Languages', trans. J.H. Woodger, in *Logic, Semantics and Metamathematics*, Oxford: Clarendon Press, 152-278.

Tarski, A. 1944. 'The Semantic Conception of Truth', *Philosophy and Phenomenological Research* 4, 341-375.

Tichý, P. 1974. 'On Popper's Definitions of Verisimilitude', *The British Journal for the Philosophy of Science* 25, 155-160.

Trigg, R. 1989. *Reality at Risk*, 2nd edition, New York-London-Toronto-Sydney-Tokyo: Harvester Wheatsheaf.

Unger, P. 1971. 'A Defense of Skepticism', *The Philosophical Review* 80, 198-219.

Van Cleve, J. 1984. 'Reliability, Justification, and Induction', *Midwest Studies in Philosophy* 9, 555-567.

Van Fraassen, B. 1980. *The Scientific Image*, Oxford: Clarendon Press.

Vollmer, G. 1987. 'On Supposed Circularities in an Empirically Oriented Epistemology', in G. Radnitzky and W.W. Bartley, III eds. *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*, La Salle Ill.: Open Court, 163-200.

Weston, T. 1992. 'Approximate Truth and Scientific Realism', *Philosophy of Science* 59, 53-74.

Williams, B. 1978. *Descartes: the Project of Pure Enquiry*, London: Pelikan.

Wittgenstein, L. 1953. *Philosophical Investigations*, Oxford: Blackwell.

Wright, C. 1992. 'On Putnam's Proof that We are not Brains-in-a-Vat', *Proceedings of the Aristotelian Society* n.s. 92, 67-94.

Zahar, E. 1984. 'The Popper-Lakatos Controversy in the Light of *Die beiden Grundprobleme der Erkenntnistheorie*', *The British Journal for the Philosophy of Science* 35, 149-171.