The Externalist Challenge

Current Issues in Theoretical Philosophy

Edited by
Richard Schantz

Vol. 2

W
DE
G

Walter de Gruyter · Berlin · New York

The Externalist Challenge

Edited by
Richard Schantz

Contents

Introduction

This book is the second of a series of three volumes dedicated to central debates in contemporary theoretical philosophy. The title of volume I is *What is Truth?*, the title of volume III will be *Prospects for Meaning*.

The debate between internalism and externalism has become a focal point of attention both in epistemology and in the philosophy of mind and language. Externalism challenges basic traditional internalist conceptions of the nature of knowledge, justification, thought and language. What is at stake, is the very form that theories in epistemology and the philosophy of mind ought to take. This volume is a collection of new essays reflecting the present state of the art concerning the exciting controversies between internalism and externalism.

The Debate between Internalism and Externalism in Epistemology

Epistemic Internalism

I will start with some comments on the internalism/externalism distinction as it arises in epistemology. Historically, from Plato to the present day, most theories of knowledge have been internalist in character. The traditional or standard analysis of knowledge holds that knowledge is justified true belief. Since, according to this analysis, justification is a necessary condition of knowledge, epistemologists have made every endeavour to develop theories of what epistemic justification or rationality consists in. The rationale for thinking that justification is necessary for knowledge is that merely accidentally true belief, beliefs held for bad reasons or no reasons at all, do not amount to knowledge. Justification has been the centre-piece of the theory of knowledge. And it is the central claim of internalism that typical epistemological questions such as "What is knowledge?", "What can we know?", "How can we be sure that our beliefs are justified?", "Can we be absolutely certain that we know anything at all?" and "Does knowing entail knowing that one knows?" arise from within a person's subjective cognitive perspective, and can only be adequately answered by referring to resources accessible to that person from that perspective.

The modern era in philosophy, with its characteristic epistemological turn, was initiated by Descartes. The most conspicuous feature of his approach to epistemological problems was its radically internalist orientation. The aim of his famous "method of doubt" is to subject all our beliefs and principles to radical scrutiny, to test them to the limit, and to accept, finally, only those beliefs that, logically speaking, cannot be doubted. By applying a series of elaborate sceptical techniques, he intends to eliminate all dubious beliefs in order to find the absolutely certain ones, those that stand up to even the most extreme sceptical attack. According to Descartes, there are two kinds of

beliefs that survive the process of systematic doubt, that are infallible, guaranteed to be true, and thus can count as knowledge: firstly, the belief that he himself exists as a thinking self, and, secondly, equally important, the beliefs he has about the specific contents of his own sensory experiences and other mental states. It is these latter beliefs that form the foundation, the empirical basis, of all other kinds of knowledge, including above all knowledge of the external world. So Descartes' foundational project is an ambitious attempt to reconstruct knowledge from the inside outwards—knowledge of our sensory experience is the starting point from which knowledge of the external world has to be inferred. His epistemological internalism comes to light in his principal assumption that only mental contents can function as the immediately accessible evidence that is requisite to provide a secure basis for knowledge. Justification is a matter of things purely internal to the subject's mind, a matter of things of which the subject can be justifiably aware of. Many philosophers still follow Descartes' lead. His resolute first-person perspective on the nature of epistemological problems has become part of our theoretical heritage.

Here is a more modern statement of the internalist point of view by Roderick Chisholm, one of the most important epistemologists of the last century:

> The usual approach to the traditional questions of theory of knowledge is properly called "internal" or "internalistic". The internalist assumes that merely by reflecting upon his own conscious state, he can formulate a set of epistemic principles that will enable him to find out, with respect to any possible belief he has, whether he is *justified* in having that belief. The epistemic principles that he formulates are principles that one may come upon and apply merely by sitting in one's armchair, so to speak, and without calling for any outside assistance. In a word, one need consider only one's state of mind.[1]

So what is required for a belief to be epistemically justified for a given person is for that person to have cognitive armchair access to evidence that supports the truth of the belief. What can serve for justification must be something like a belief or an experience, something of which the person has a cognitive grasp. The requirement that all justifying factors must be purely internally available can be interpreted in at least two ways: a strong position requires that the subject actually have the knowledge or justifiable belief that the justifier obtains, whereas a weaker version requires only that the subject be capable of acquiring this knowledge by appropriate focussing of attention, without any change of position, new information, and so on. The central idea is that by reflection or introspection alone we can determine whether a certain belief is justified or not.

It is crucially important to distinguish between access to the justifier and access to the epistemic efficacy of the justifier. In order for a person to know or justifiably believe that *p*, it is not enough that she has access to other justified beliefs or other justifying factors, such as sensory experiences, providing her with adequate evidence for the belief that *p*. She must also know or justifiably believe that these other justifying factors are indeed adequate reasons or grounds for her belief that *p*. But it is far from obvious that the satisfaction of such severe requirements can be ascertained just on re-

---

1   Chisholm 1989, 76

flection. This distinction opens up the possibility of advocating a mixed position, one demanding the cognitive accessibility of the ground on which a belief is based but not of its evidential adequacy.

No doubt, internalism possesses a strong intuitive plausibility. How could something that does not fall within a person's ken, something to which perhaps she has no access at all, give her a reason for one of her beliefs? Internalists typically lay stress on having reasons for our beliefs, and this is typically interpreted as being able to use those reasons as premises in arguments, or to appeal to them when questioned. This is not something one can be unaware of. It has been argued, however, that part of internalism's intuitive appeal rests on a confusion between having reasons and giving reasons or, in other words, between the state or property of being justified in believing, and the activity of justifying the belief, of showing it to be justified. If one neglects this important distinction, one will be inclined to suppose that being justified entails the ability of showing that one is justified, that is to say, of producing a justifying argument. Now, of course, to successfully carry out the activity of justifying a belief one must appeal to other beliefs or other conceptually contentful states. But from the fact that we can *justify* a belief only by showing that it sustains appropriate relations to other beliefs, it does not follow that a belief can *be justified* only by its relations to other beliefs. We would have only few justified beliefs, if their justification depended on having engaged in the practice of establishing their credentials.[2]

Epistemic Deontologism

The main motivation for the accessibility constraint of internalism derives from a deontological conception of justification, which has its origin in the works of Descartes and Locke. Descartes, in the fourth of his *Meditations*, explains error as the wrong use of the will in judging. He argues that the will is not limited by the perceptions of the understanding, so that we are free to choose, by a bare act of the will, whether or not to accept a proposition. The policy he puts forward is that we should only believe those propositions whose truth we clearly and distinctly perceive. Erroneous beliefs result when the limitless will exceeds the limits of the understanding, that is, when one attaches assent to a proposition p in the face of the recognition that one's evidence for p is less than adequate. We are fully responsible for these misuses of our intellectual freedom and may, therefore, justly be blamed. The virtuous course in these cases is to suspend judgement on p. Similarly, Locke, in his *An Essay concerning Human Understanding*, speaks of epistemic duties that we, as "rational creatures", have to meet and for the violation of which we are held accountable. In his view, our primary epistemic goal is to believe what is true and not believe what is false. This goal imposes on us the duty to seek the truth to the best of our intellectual faculties.

---

2   See Alston 1989

In analogy to ethical deontologism which maintains that we have duties to act in certain ways and that we are responsible for our actions, epistemic deontologism maintains that we, as intellectual beings, have duties to believe on the basis of adequate evidence and that we are responsible for our beliefs. According to this conception, a person is justified in believing that p if and only if in holding the belief she satisfies, or does not violate, any intellectual duties or obligations. Epistemic justification is a normative notion, pertaining to what we should or should not believe. So it seems that the focus has to be on our epistemic responsibility, on whether or not a belief has been responsibly formed or is responsibly held. We are blameworthy if we do not live up to appropriate standards of epistemic conduct, if, for example, we base our judgment on misleading evidence and so acquire commitments that we should not hold. From the standpoint of epistemic deontologism, a justified belief is a belief which, by reference to our epistemic duties, it is permissible to hold; we cannot rightly be reproached for holding it.[3] But, so internalists argue, since a belief is permissible only if we have a justification for it, we must have access to the conditions that determine whether our beliefs are justified. Otherwise a person could do her epistemic best, could do as demanded of her, and hence be praiseworthy but still hold unjustified beliefs.

Deontological conceptions of justification seem to presuppose a kind of doxastic voluntarism. As John Pollock pointedly puts it: "I have taken the fundamental problem of epistemology to be that of deciding what to believe."[4] And indeed, in order for talk of epistemic duties, responsibilities and permissions to make sense, we must have direct voluntary control over our beliefs. After all, I cannot have a positive duty to believe a proposition or a negative duty to refrain from believing it if it is not up to me to decide as to whether to believe it or not. We often hear that what a person ought to do, she can do. The dependence on doxastic voluntarism, however, brings to light a serious problem for epistemic deontologism. The trouble with doxastic voluntarism is that, as has been pointed out by William Alston and Alvin Plantinga above all, our beliefs are not generally under our direct voluntary control.[5]

Although the relationship between the will and belief is a complex affair, I think it is fair to say that, at least for the most part, we do not choose our beliefs. We simply cannot believe or refrain from believing at will. Think of beliefs resulting from perception. When I see a cat on the sofa, and have no reason to suppose that the conditions of perception are abnormal, I am simply not free as to whether to believe that there is a cat on the sofa. I will form that belief, whether I like it or not. Deliberate choice does not play any role; a basic act of will is typically not involved. Rather, acquisitions of perceptual belief exhibit a characteristic involuntariness and passivity; they are events in which an external state of affairs forces itself on us. Or reflect on introspective beliefs. When I believe that I have a toothache, I cannot, by performing an act of the will, just choose to stop believing this. As long as I have the toothache, I will not succeed in getting rid of this belief, no matter how hard I try. Finally, a consideration of beliefs based

---

3   See Ginet 1975
4   Pollock 1986, 10
5   Alston 1989; Plantinga 1993a

on reasoning also confirms the verdict that we cannot simply obtain beliefs through consciously willing to have them. If we accept the premises of a valid deductive argument, we must, are forced to, accept its conclusion too—regardless of whether we wish it to be true. None of these beliefs can be subject to epistemic duty.

What has been criticized is the view that, generally, our beliefs are directly subject to effective voluntary control. This, however, is not to deny that we have some kind of indirect control over our beliefs or world-views. For we have voluntary control over many activities that influence their generation. Many of our beliefs are the results of explicit decisions as to how to conduct inquiry. It is up to us whether we investigate matters thoroughly, try to reason carefully, pay attention to evidence on both sides of an issue, seek the advice of other people, and so on. And we even have some kind of control over things that affect perceptual beliefs; even perception of the world involves an active element. We can try to make sure that the conditions of perception are optimal, direct our perceptual systems, focus on certain aspects, abstract from others, and the like. In these indirect, derivative senses, we are responsible for the beliefs we finally arrive at. But such an indirect control over our beliefs is not what the epistemic deontologist needs in order to support an internalist conception of justification.

Arguments against Epistemic Internalism

The main difficulties of epistemic internalism can be traced back to the accessibility requirement. Its advocates seem to overestimate the cognitive powers of armchair reflection. We have seen that Chisholm claims that just by carefully thinking about it, we are capable of finding out what provides a justification for what. How questionable this claim is, becomes obvious when we look at the alternative non-deontological account of justification that is usually called a "truth-conductivity" conception. According to this widely held conception, justification is essentially or internally related to the cognitive aim of truth, to the aim of maximizing truth and minimizing falsity. Thus a person is justified in believing that p only if she is in a strong position to attain the truth, only if she has an adequate reason to believe that p is true. An adequate reason is, on this conception, a kind of truth indicator, something that makes it evidentially at least very probable that the belief that p be true. On this account, a belief's being justified is a normative status, but the normativity in question has nothing to do with responsibility, duty, or permission. It is simply a matter of being appraised as good or bad, favorable or unfavorable, with respect to the basic epistemic end of seeking true beliefs and avoiding false ones.

Here is how Lawrence BonJour, interestingly an unwavering defender of internalism, brilliantly formulates and defends the truth-conductivity conception of justification:

> And, if our standards of epistemic justification are appropriately chosen, bringing it about that our beliefs are epistemically justified will also tend to bring it about ... that they are true. If epistemic justification were not conducive to truth in this way, if finding epistemically justified beliefs did not substantially increase the likelihood of finding true ones, epistemic justification would be irrelevant to our main cognitive goal and of dubious worth. It is only if we have some reason to think that epistemic justification constitutes a path to truth that

we as cognitive human beings have any motive for preferring epistemically justified beliefs
to epistemically unjustified ones. [6]

If, however, conditions of justification have to be such that the fulfilment of those con-
ditions guarantee that the belief in question is at least likely to be true, then the idea
that we can determine by mere reflection whether certain conditions justify a given
belief seems to lose its original intuitive appeal. According to familiar versions of coher-
entism, for example, epistemic justification is a holistic notion; an individual belief is
justified only if it coheres, or harmonizes, in an appropriate way with a person's belief
system, that is, with the totality of her beliefs. Since it is psychologically impossible
for an agent exactly to survey the set of all of her beliefs, even coherentism seems to
be unable to meet the accessibility constraint. The central problem for epistemic de-
ontologism is that a belief might enjoy deontological justification without being based
on a truth-conducive ground.

In face of these difficulties it is small wonder that more modest versions of internal-
ism have been proposed, ones which, while still holding that justification supervenes
upon internal psychological conditions, drop the further requirement that these condi-
tions must be reflectively accessible.[7] Thus, modest internalism opens up the possibility
that the facts on which the justification of a belief depends include, e.g., neural events
proximate to the belief in question. But whether or not internal facts of these sorts
obtain can hardly be determined by introspection alone. So this is indeed a modest or
weak version of internalism—more defensible, surely, than accessibility internalism but
without the latter's original intuitive appeal.

Epistemic Externalism

Pure externalist views deny that justification, at least in the customary sense, is a neces-
sary condition for knowledge. Thus, they reject the traditional justified true belief ac-
count of knowledge. Advocates of externalist accounts of knowledge draw attention to
features of the world other than the subject's reasons for belief. They urge that not all
true beliefs need to be supported by reasons, classically conceived, in order to qualify
as knowledge. Hence their view looks outward to the facts the beliefs are about, rather
than inward to the reasons a subject has for holding her beliefs. David Armstrong,
who introduced the term "externalism" as a terminus technicus in epistemology in his
pioneering book *Belief, Truth and Knowledge* of 1973, characterizes this position in the
following way:

> According to "Externalist" accounts of non-inferential knowledge, what makes a true non-
> inferential belief a case of *knowledge* is some natural relation which holds between the belief-
> state, Bap, and the situation which makes the belief true. It is a matter of a certain relation
> holding between the believer and the world. It is important to notice that [...] Externalist

---

6   BonJour 1985, 8
7   Feldman and Conee 1985

theories are regularly developed as theories of the nature of knowledge *generally* and not simply as theories of non-inferential knowledge.[8]

The core of the non-justificational analysis of knowledge Armstrong here proposes is that knowledge merely requires some sort of natural relation between a person's belief and the fact which makes it true. A good example is Alvin Goldman's early causal theory of knowledge, developed primarily as a response to the Gettier problems, according to which a person's true belief that p counts as knowledge just in case it has the right sort of causal connection to the fact that p.[9] The sorts that are right, or appropriate, comprise perception, memory, good reasoning and various other kinds of causal chains, and combinations of these. In Gettier examples, a person has a justified belief that is only accidentally true. The causal theory explains the lack of knowledge in these cases, because in them the basic requirement that there be an appropriate causal connection between the fact that p and the person's belief that p is not satisfied. Standard cases of perception provide a useful model of the right sort of causation. Fred Dretske, refining earlier accounts by explicating the causal relation in terms of the reception of information from a source, developed an influential information-theoretical analysis of perceptual knowledge on which a subject S knows that x is F just in case her belief is caused or causally sustained by the information that x is F received from the source x by S.[10]

Goldman, originally, regarded the causal criterion as both a necessary and sufficient condition for knowledge. He did not just suggest supplementing the justification condition with his causal condition. For Goldman, then, the justification condition had to be replaced with an external standard—an appropriate causal relation to the external world. Later, Goldman came to advocate a related theory, the reliability theory of knowledge, which Armstrong had favoured from the beginning.[11] Remarkably, it was Frank Ramsey who first defended a reliability theory of knowledge.[12] Theories of this kind abandon the prerequisite of earlier, and simpler, causal theories that there be an actual causal chain linking the fact represented in the belief with the belief itself. Instead, proponents of the reliability account typically hold that a belief qualifies as knowledge just in case it is both true and empirically reliable, that is, just in case the belief is produced in a way that makes it objectively probable that the belief be true. There are two main types of reliabilism: reliable indicator theories and reliable process theories. According to theories of the first type, a belief counts as knowledge when it is a reliable indicator of its truth. According to theories of the second type, a belief counts as knowledge just in case it is produced, or is sustained, by a generally reliable process, that is, one that leads to a high proportion of true beliefs.[13]

Armstrong's account belongs to the first type. He argues that the tight connection between a person's belief and the fact making it true which knowledge demands should be construed as a nomological connection. To illustrate his point, he draws an analogy

---

8 Armstrong 1973, 157
9 Goldman 1967
10 Dretske 1981
11 Goldman 1979; Armstrong 1973
12 Ramsey 1931
13 Goldman 1976, 1979; see also Talbott 1990

between a thermometer which reliably indicates the temperature and a belief which reliably indicates the truth. His claim is that a belief about a perceived object x, that it is F, is non-inferential knowledge if and only if the belief is a completely reliable sign that x is F. This depends on properties of the believer—e.g. that her senses and belief-forming mechanisms are functioning properly in a suitable context—such that lawlike connections of nature determine that, for any subject S and object x, if S has those properties and believes that x is F, then x is F.

Similarly, Dretske maintains that the reasons we have for holding our beliefs should be nomologically connected to their truth, that is, it should be a law of nature that a person having such reasons for her belief will have a true belief. Since lawlike statements, in contrast to statements of mere uniformity, support subjunctive or counterfactual conditionals, the central idea of Dretske's account can also be cast in the subjunctive mode. Arguing that the possession of "conclusive reasons" to believe is both a necessary and a sufficient condition for knowledge, he goes on to characterize conclusive reasons by using a subjunctive construction: R is a conclusive reason for p if and only if R would not be the case unless p were the case.[14]

A related proposal was made by Robert Nozick who claims that knowledge is true belief that "tracks the truth". Nozick also uses various subjunctive or counterfactual conditionals to explicate the specific sense in which knowledge is belief that co-varies with the facts and is thereby distinguished from accidentally true belief. He suggests, firstly, a "variation" condition: if p were not true, then S would not believe that p. But he argues that, secondly, a further clause, an "adherence" condition, must be added: if, in slightly different conditions, p were still true, then S would still believe that p. A variant of the subjunctive account of knowledge, the theory of relevant alternatives, holds that S knows that p only if there is no relevant alternative situation in which p is false but S would still believe that p. So in order to know that p, S must be able to discriminate the actual situation in which p is true from all relevant counterfactual situations in which p would be false.[15]

Goldman, later, came to adopt a reliable process theory. In contrast to Armstrong's and other externalists' estimation of reliability theories, though, Goldman now introduced an important new perspective. Instead of rejecting the traditional view that knowledge demands justified true belief, he now argues for the alternative that we should think of reliability not as replacing justification but, rather, as providing a new analysis, a naturalistic reduction, of what justification amounts to. Since, for him, justification is by definition that which has to be added to true belief to convert it into knowledge, justified belief just is reliably produced belief. Similarly, Marshall Swain, presenting a version of reliabilism in terms of probability, holds that a belief is justified when it is based on reasons that are reliable indicators of the truth.[16]

A crucial feature of reliabilism is that it does not require that the person have any sort of cognitive access to the fact that the belief-producing processes are reliable in

---

14 Dretske 1971
15 Goldman 1976
16 See Swain 1981

order for her belief to be justified. All that matters is that the actual causal process by means of which the belief in question is generated is in fact reliable, whether or not the believer, or anyone else, has at the time, or indeed ever, any beliefs about the fact of reliability That is why reliabilism is a form—indeed the most influential form—of externalism: it makes the epistemic status of a belief depend on external factors, factors concerning the believer's environment, history or social context to which she may have no cognitive access at all.

Against this background, externalists tend to warn of a level confusion, the confusion between knowing, or being justified in believing, that p—first-level knowledge or justification—and knowing, or being justified in believing, that one knows, or is justified in believing that p—second-level knowledge or justification. They insist that unreflective first-level knowledge does not require that the believer knows anything about her own reliability. She knows just in case her belief is *de facto* truth-reliable. Of course, if she has no beliefs on the reliability of her cognitive mechanism, she lacks something. And what she lacks, is second-level knowledge or justification: she does not know, or justifiably believe, that her belief is justified or amounts to knowledge. No wonder then that externalists generally dismiss the KK-thesis, the thesis saying that if a person knows that p, then she knows that she knows that p, which was tacitly or explicitly accepted by many philosophers since Plato and Aristotle.

An offshoot of reliabilism is Alvin Plantinga's comprehensive theory of warrant and proper function.[17] Plantinga emphatically repudiates epistemic internalism and with it the whole idea of justification, replacing it with the concept of warrant which he too characterizes as the quality that when added to true belief yields knowledge. Purporting to go beyond reliabilism, this theory does not chiefly focus on reliable processes but rather on cognitive faculties that function according to a design plan in a suitable environment. A belief is warranted only if it is produced by a properly functioning faculty in a cognitive environment that is fitting for this kind of cognitive faculty— the environment must be the one or like the one for which it is designed. Either God or natural selection designed us to function successfully in our surroundings. On the other hand, Plantinga critically examines and rejects, sophisticated as they may be, the leading naturalist accounts of proper function that have been proposed. He argues that the notion of function demands the notion of design, and design in turn demands something non-natural. Hence, ultimately, his whole intricate theory calls for some theistic or supernatural underpinning. The design plan turns out to be the divine plan. Plantinga himself confesses to be a naturalist in epistemology, but a supernaturalist in anthropology.[18]

Yet another kind of reliabilism is presented by virtue epistemology and its attempt to make epistemic character traits, rather than epistemic principles or duties, the centre of epistemology. The leading figure of this approach to date is Ernest Sosa. The basic idea of his "virtue perspectivism" is that justified belief is belief that arises from the proper functioning of our intellectual virtues in an appropriate environment, and

---

17 Plantinga 1993a, 1993b
18 Plantinga 1993b, 46

not from intellectual vices.[19] An intellectual virtue is a competence or faculty or pow-
er of the mind to believe true propositions and to avoid believing false ones. So the
virtues are truth-conducive, whilst the vices are inimical to truth. Human epistemic
virtues are reliable cognitive mechanisms such as perception, memory, introspection,
deduction and induction, whereas vices include wishful thinking, hallucination and
clairvoyance. In contrast to generic reliabilism, virtue perspectivism restricts the class
of reliable processes that are epistemically significant to those processes that have their
source in an epistemic virtue, in a stable disposition for belief acquisition. Moreover,
Sosa distinguishes between "animal knowledge" and "reflective knowledge". The latter
higher sort of knowledge, the sort of knowledge traditionally desired, requires not just
a virtuously functioning faculty, but, additionally, that the subject enjoys an appropri-
ate epistemic perspective on her first-level beliefs, that she grasps the reliability of her
faculties. Finally, virtue perspectivism differs from Plantinga's theory by eschewing the
contentious notion of a design plan for our epistemic faculties.

Advantages of Epistemic Externalism

Externalism holds many attractions for epistemologists. First of all, it offers a scientifi-
cally respectable account of knowledge. Many of its proponents contend that external-
ism is a form of epistemology naturalized and so follow Willard Van Quine's advice to
look upon epistemology as an enterprise within the framework of natural science, that
is, to transform epistemology into a chapter of empirical psychology. The prime goal
of naturalized epistemology is to answer the question of how we acquire our overall
theory of the world. By adopting a genetic approach, Quine and his followers seek to
provide a factual account of the link between sensory evidence and theory. In this un-
dertaking, cognitive agents are viewed as natural phenomena, and so the focus of the
scientific investigation of the acquisition of knowledge is on the causal relations between
their sensory input and their theoretical output. Moreover, externalism is naturalistic
in that it tries to reduce normative epistemic properties to natural properties and rela-
tions, the paradigm case being the attempted reduction of justification to reliability.
And indeed the study of psychological processes and mechanisms of belief formation
and alteration seems to be a descriptive and not a normative affair.

   Secondly, externalism seems to be capable of protecting knowledge against the threats
of radical scepticism. Instead of setting extremely high standards for knowledge, exter-
nalism proposes more moderate constraints. True beliefs that arise via a reliable method
are knowledge because of the way they arise. The subject need not, in addition, know
that they have arisen in that way rather than in the ways radical scepticism suggests. It
is a result of this modesty that we can quite plausibly attribute simple factual knowl-
edge to young children and even to animals, as we indeed usually do, without having
to assume that they are in possession of sophisticated knowledge or beliefs about the
epistemic status of their beliefs.

---

19 Sosa 1991

Arguments against Epistemic Externalism

No doubt externalism has many advantages. Nevertheless, there are also several serious problems confronting this position. The main objections to externalism are based on deeply rooted internalist intuitions. Many internalist critics are fond of pointing out that externalism marks a significant break from the modern epistemological tradition originating from Descartes. So we find BonJour saying:

> When viewed from the general standpoint of the western epistemological tradition, externalism represents a very radical departure. It seems safe to say that until very recent times, no serious philosopher would have dreamed of suggesting that a person's beliefs might be epistemically justified simply in virtue of facts or relations that were external to his subjective conception. Descartes, for example, would surely have been quite unimpressed by the suggestion that his problematic beliefs about the external world were justified if only they were in fact reliably related to the world—whether or not he had any reason for thinking this to be so. Clearly his conception, and that of generations of philosophers who followed, was that such a relation could play a justificatory role only if the believer possessed adequate reason for thinking that it obtained.[20]

With this general point in mind, BonJour and others have elaborated their criticisms of externalism by arguing that the satisfaction of externalist conditions is neither necessary nor sufficient for epistemic justification. The first objection challenges the necessity of the reliability requirement by considering demon worlds, worlds controlled by an evil omnipotent demon who produces, in the denizens of such a world, patterns of sensory experience exactly like our own, even though, in their case, the resulting perceptual beliefs are systematically false.[21] Despite the unreliability of their belief-generating processes, internalists quite plausibly claim that the believers in a demon world would be justified in their beliefs, just as much so as we are in ours. After all, they have the same reasons for belief as we do, and their whole epistemic situation may well be subjectively indistinguishable from our own. So the conclusion seems to be that reliability cannot be necessary for justification.

Some externalists have responded by simply rebuffing the intuitions adduced by their internalist critics. Their hardline diagnosis is that these intuitions are misleading and in the end merely reflect the persistent internalist strand of the epistemological tradition. Externalists typically are supporters of the truth-conducivity view of justification. And if this view is correct, if justification indeed has some essential relation to the goal of truth, then the beliefs of the inhabitants of the demon world just are not justified—even though they may have done their best to bring their intellectual house in order.

Other reliabilists, however, have responded by modifying and refining their theories. One kind of reliabilism, normal-worlds reliabilism, rejects a crucial assumption upon which the counterexample rests, the assumption namely, that the reliability of a cognitive process is to be assessed by its performance in the world that the believer in question inhabits.[22] Instead, so their alternative proposal, reliability should be assessed in normal

---

20   BonJour 1985, 56
21   Foley 1987
22   Goldman 1986

possible worlds, that is, in worlds consistent with our general beliefs about the actual world, about the kinds of objects and events that occur in it. Since the cognitive processes deployed by the victims of an evil demon apparently would be reliable in normal worlds, the reliabilist can now agree that the use of these same processes will generate justified beliefs even in a demon world. Thus we get the intuitively right result.

The second, complementary objection to externalism tries to show that reliability is not sufficient for justification. Imagine a situation in which a person forms beliefs in a highly reliable way, that is, in which the requisite externalist conditions are satisfied, but in which she has no reason to regard her mode of belief formation as reliable and perhaps good reason to regard it as unreliable. Thus suppose that there is some person with a reliable clairvoyance capacity who forms beliefs about distant events on this basis, and these beliefs are usually correct.[23] Suppose further that this person has no evidence to think that she possesses such a cognitive capacity, or perhaps she even has strong evidence to the contrary. Internalists' intuitive claim is that, although the reliabilist conditions are clearly satisfied, this person is highly irrational or irresponsible and hence not epistemically justified in confidently accepting her clairvoyantly formed beliefs.

While, again, several externalists have simply dismissed the intuition as misguided, others became convinced by this sort of case that a further strengthening of the basic reliabilist constraints is needed. So some externalists have introduced a "non-undermining" condition, a condition to the effect that, roughly, the subject neither believe nor have good reasons for believing that the generating cognitive process is unreliable.[24] So if the subject has adequate evidence against her possession of reliable clairvoyance powers, then the beliefs she arrives at on this basis would, according to this modified version of reliabilism, be unjustified. The new non-undermining condition would be violated.

Yet another kind of problem facing reliabilism is the so-called "generality problem".[25] Reliabilism holds that a belief is justified if the cognitive process that causes it is reliable. But processes can be individuated in many ways. Whenever a particular belief is produced, the process token that induces it can be viewed as an instance of numerous process types of differing generality. The serious problem for reliabilism is that these different process types have different truth ratios, and hence different degrees of reliability. So which of these many types should be used in fixing reliability? Should the relevant types be construed broadly or narrowly? If the types are characterized too broadly, many clearly unjustified beliefs will have to be regarded as justified. Perceptual belief formation, for example, is usually deemed to be a reliable process, but it brings about unjustified beliefs as well as justified ones. Obviously, visual perceptions of objects at close range under optimal conditions are more reliable than visual perceptions of distant objects under substandard conditions. And even if the conditions of perception are optimal, some of the beliefs formed in such conditions, typically the more general

---

23   BonJour 1985
24   Goldman 1986
25   Feldman 1985

ones, are less likely to be mistaken than others. If, on the other hand, the types are characterized too narrowly, including detailed features of some specific situation, they may turn out to have only one actual instance, namely, the process token in question. In this "single case" scenario, the process type will be either 100 percent reliable or 100 percent unreliable, depending on whether the belief is true or false. Consequently, the belief is justified if and only if it is true, which is not a welcome result.

The problem of the single case can be avoided, however, if general reliability is understood as a propensity, the propensity to produce a high truth ratio of beliefs, rather than as a strict frequency. Since, according to the propensity construal, actual track record is not sufficient, the additional problem is raised of adequately specifying the range of possible uses of the process. It seems reasonable, for various reasons, only to take into account the performance of the process in nearby possible worlds. With these plausible answers to the problem of the single case and to the related problem of the range of reliability, the way seems paved, finally, for a narrow principle of type individuation. For, plainly, it is narrow types that are needed to differentiate between those cognitive processes that do lead to knowledge and those that do not. Against this rather complex background, full of twists and turns, Goldman arrives at his well-considered solution of the generality problem. He advances the conjecture that the decisive psychological process type is the narrowest type that is causally operative in eliciting the belief token in question.[26]

Externalism versus Internalism about Content

Many of our mental states, processes and events, such as our beliefs, desires, perceptions and memories, are semantically evaluable. In contrast to stars, tables, or tomatoes, mental states have representational content or exhibit the feature of intentionality. When we believe or desire or perceive, we always believe or desire or perceive something. So what is characteristic of states with content is that they are directed at, or about, objects and states of affairs in the world. To be in a contentful mental state is to represent the world as being a certain way. In believing, e.g., we believe about certain things that they have certain properties or stand in certain relations. Therefore, beliefs are true or false depending on whether things are the way they are believed to be. Of course, mental states can fail to refer. Nevertheless, a non-referring mental state still has a satisfaction condition, a condition, specified by its content, which its purported referent, and the properties of its referent, would have to satisfy for the state to be correct. Representational states are what they are in virtue of their content, in virtue of how they represent things as being. Consequently, in our customary cognitive practice we find it appropriate to individuate such states, that is, to distinguish them from another, by their different contents.

---

26 Goldman 1986, 49-51

But what is the nature of content, the nature of the things that we think or believe or desire? Answering this question involves taking sides in the fundamental debate between internalism, also often called "individualism", and externalism in the philosophy of mind. Externalism and internalism proffer different conceptions of the relations between mind and world. They are typically understood as theses about the individuation of mental states, that is, as theses about the existence and identity conditions of mental states. Externalism claims that the existence and identity of entities outside of the thinker herself determine the existence and identity of her mental states. So, according to externalism, a belief, say, is intrinsically dependent upon the environment of the believer. There is no way to characterize the beliefs a person has save by characterizing the disposition of objects and properties in the external world that the belief is about. Thus the surrounding world is taken to be constitutive of the very nature of beliefs; after all, the world fixes what they are. There is no sharp metaphysical boundary separating mind and world.

Content Internalism

Internalism offers a quite different picture of the relationship between mind and world. Its defenders assert just what externalists dispute, namely that the contents of the mind can be described in ways that do not require essential reference to how things are in the world outside. They can see no principal reason to embrace their opponents' view that mental content is essentially dependent on the physical or social environment in which we are situated. In sharp contrast to externalists, internalists maintain that the content of mental states is determined only by the internal, non-relational properties of the subject's mind or brain. For them, the mind is largely autonomous with respect to the world—not deeply penetrated by it. Of course, internalists, typically, do not deny that our beliefs are about things and states of affairs in the natural world. They do not reject the very idea of intentionality. Rather, what they insist upon, is that a belief could exist and have the very same content that it actually has even if its alleged object itself did not exist. So, according to internalism, mental content cannot have a relational nature.

The traditional view of content has been strongly internalist in character. Descartes' various sceptical arguments all rest on the assumption that our knowledge of the content of our sensory experiences, and of the content of the beliefs resulting from them, is logically independent of any knowledge about the external world. Since there is no deductive bridge linking the way things appear to us and the way they really are, we cannot exclude the possibility that the world is very different from the way we take it to be. Hence, all our beliefs about the external world might be massively false. But, on the Cartesian picture, even if our beliefs were massively false in this way, even if a world outside our minds did not exist at all, we could still have the same sensory experiences and the same beliefs that we actually have. We could have exactly these experiences and exactly these beliefs even if we were brains in a vat—provided, of course, something elicited in our brains the same internal states that we in fact enjoy. So, on the Cartesian view, the content of an experience, or of a belief, is an internal property

of the subject. The individuation of mental states is completely independent of any external objects or states of affairs.

Descartes is a metaphysical dualist; a sharp boundary separates the self-sufficient inner realm of the mind from the external realm of material substances. His metaphysical dualism of substances is accompanied by the epistemological claim that there is a fundamental asymmetry between our ways of knowing about our own minds and about the external world, including other persons' minds. We have a special direct cognitive access to what passes in our minds which contrasts sharply with our indirect ways of acquiring knowledge about the rest of the world. On the Cartesian conception, we have, as it is often put, first-person authority with respect to the contents of our minds. Descartes' firm conviction that we are authorities about our own minds is grounded in his doctrine of self-intimation, the doctrine that our whole mental life is transparent to our introspective awareness. Moreover, he is convinced that introspective awareness is infallible.

The Cartesian philosophy of mind, by setting the world dualistically over against our conceptual scheme, presupposes that mental individuation can indeed be explained without recourse to worldly conditions. It presupposes just the existence of thinkers and their thoughts. Descartes is the arch-internalist about the nature of mental content. After him, Locke, Leibniz, and Hume, despite the differences between their specific theories, shared most of the Cartesian internalist assumptions about the constitution of the content of mental representations. Gottlob Frege seems to be another internalist about content. He says, for example:

> Let us just imagine that we have convinced ourselves, contrary to our former opinion, that the name Odysseus, as it occurs in the Odyssey does designate a man after all. Would this mean that sentences containing the name "Odysseus" expressed different thoughts? I think not. The thoughts would strictly remain the same; they would only be transposed from the realm of fiction to that of truth. So the object designated by a proper name seems to be quite inessential to the thought-content of a sentence which contains it.[27]

On Frege's semantic theory, there is a special Platonic realm of sense mediating between language and the realm of reference. A thought, as he uses the notion, is the sense expressed by a complete assertoric sentence. He characterizes sense as that property of an expression that has to be "grasped" in order to understand it. And though the sense of an expression is also described as the "mode of presentation" of its reference, his theory allows for the possibility that an expression may have a sense but lack any reference. Frege is asserting that our grasp of the sense of the name "Odysseus" is independent of whether or not it has a reference. Therefore understanding an expression, consisting as it does just in the grasping of entities of the realm of sense, occurs in splendid metaphysical isolation from the external physical world. Thinking, for Frege, is a mental activity in which we, in some unexplained manner, grasp, or apprehend, peculiar Platonic thoughts—objects which are mind-independent, objective and public, but nonetheless neither in space and time nor perceptible. Frege also believes in the trans-

---

27 Frege 1919, 191

parency of sense. By grasping the sense of an expression we know everything there is to be known about its internal properties; a sense cannot have any properties that are undetectable by us. This is reflected in his criterion of identity, according to which a sentence S expresses a different thought than a sentence S' just in case it is possible for someone to understand both sentences while taking one to be true and the other to be false. We are capable of determining, by introspection alone, whether or not two sentences express the same or different thoughts. Concerning this criterion, the actual truth values of the sentences—their references, on Frege's view—are entirely irrelevant. In Frege's philosophy, a Cartesian view of the mind coalesces with a Platonic view of its contents.

In contrast to Descartes' and Frege's dualism, many modern partisans of internalism rather tend to sympathize with a materialist metaphysics. If they believe at all that mental states and processes possess semantic properties, if they do not flirt with one of the various variants of eliminative materialism, then they are liable to believe that these properties supervene upon inner, non-relational features of the brain or body. John Searle, for example, says:

> If I were a brain in a vat I could have exactly the same mental states I have now; it's just that most of them would be false. ... The operation of the brain is causally sufficient for intentionality. It is the operation of the brain and not the impact of the outside world that matters for the content of our mental states.[28]

Searle urges that the content of mental states is determined by operations of the brain. The psychology of an organism depends on internal physical facts; it is, in this sense, autonomous. Remarkably, in his pugnacious defense of internalism he goes so far as to claim that each of us is precisely a brain in a vat. The vat, he tells us, is a skull, and the messages from without reach us only by impacting on the brain and the rest of our central nervous system.[29]

Arguments for Content Externalism

The main arguments against content internalism and for its adversary, content externalism, derive from recent developments in the philosophy of language, associated primarily with the writings of Hilary Putnam, Saul Kripke, David Kaplan and Tyler Burge.

Presumably the most influential argument for externalism is Hilary Putnam's celebrated Twin-Earth thought experiment. Imagine that somewhere in the galaxy there is a planet, Twin Earth. Twin Earth is in almost all respects qualitative identical with Earth. Suppose that each Earthling has a doppelgänger on Twin Earth who is molecule for molecule identical with the Earthling. Moreover, our duplicates speak a language which is phonologically and syntactically the same as English. The only relevant difference between Earth and Twin Earth is that the liquid that plays the role of water

---

28   Searle 1980, 452
29   Searle 1983, 230

on Twin Earth, though phenomenologically indistinguishable from water, is not really water, that is, H_2O, but a different chemical composition, XYZ. Hence, when Oscar uses the term "water" on Earth, he refers to H_2O, that is, to real water. When, however, his doppelgänger Twin-Oscar uses the term "water" on Twin Earth, he does not refer to H_2O, but to XYZ. So what Oscar and Twin-Oscar refer to, and what they think about, is individuatively dependent on their actual environment—on the different natural kinds with which they are causally interacting. Suppose further that Oscar and Twin-Oscar are ignorant of the chemical composition of substances on their planets. Then, since they are exactly alike in every intrinsic physical respect—in the states of their brains and bodies, in their behavioural dispositions, and in their proximate patterns stimulations—this difference in reference, in extension, cannot be explained by any internal properties of them. The reference is different because the natural kind is different. And since reference is a component of meaning, Putnam concludes, memorably, that "'meanings' just ain't in the head".[30] They cannot be there because meaning depends on reference, and reference is a relation between a person and an external entity.

Other philosophers soon pointed out that externalism in the philosophy of language has to be paralleled by externalism in the philosophy of mind.[31] Meanwhile Putnam himself is convinced that not only the meanings of words but the contents of the propositional attitudes as well are determined by facts about the physical environment and, moreover, by facts about the linguistic community to which the speakers belong. Mental content too is relational in nature. This externalist turn is what in the end motivated Putnam's rejection of his former functionalist theory of the mind. If propositional attitudes aren't inner states, the nature of which can be characterized independently of external factors, then the core thesis of functionalism, the thesis that mental states are algorithmic states in our heads, cannot be right. Functionalism ultimately fails because it embodies an internalist view of the mind. It fails, in other words, because the mind is not in the head either. Rather, Putnam claims, the mind is a "structured system of object-involving abilities".[32]

Burge conducts another thought experiment designed to show that external factors enter into the determination of the meanings of our words and the contents of our thoughts.[33] Imagine an English speaker who, although he has a large number of true beliefs about arthritis, does not know that arthritis is confined to the joints. He thinks falsely that this inflammation can occur in the bones as well as in the joints. Noticing an ailment in his thigh, which is symptomatically similar to the disease in his joints, he says to his doctor, "I have arthritis in my thigh". According to Burge, he expresses, by this utterance, his false belief that he has arthritis in his thigh. When the doctor corrects his misconception, the patient relinquishes the belief. But let us now imagine a counterfactual situation in which the patient is neurophysiologically the same, but in which his social environment differs in just one respect: the word "arthritis" applies to

---

30 Putnam 1975, 227
31 McGinn 1977
32 Putnam 1992, 356
33 Burge 1979

inflammations of the bones as well as of the joints. Thus the patient's use of "arthritis" is correct in his counterfactual linguistic community. In this imagined situation, the patient would be expressing a true belief by the sentence "I have arthritis in my thigh", not a false one. So what he believes in the counterfactual case is different from what he actually believes. Yet there is no internal difference. The patient's entire physical and non-intentional mental histories remain constant. The difference in his mental contents stems from without—from differences in his social environment.

It should be pointed out that, whereas Putnam's argument applies to natural kind terms, Burge's style of argument has a much wider application. Instead of "arthritis", various other sorts of words—artefact terms, adjectives, verbs or abstract nouns —could have been used to accentuate the distinctive role that one's linguistic community plays in fixing the meanings of one's words and the contents of one's thoughts.

The relational or world-involving nature of representational content can also be brought to light by considering perceptual and indexical thoughts, thoughts whose objects are determined in a contextually sensitive way. Unlike general beliefs, genuine singular beliefs seem to require some psychologically relevant, often causal, relation between the believer and her objects; sometimes even a direct perceptual contact to items in the world, some face-to-face encounter with them, seems to be demanded. Hence these beliefs seem to essentially concern the particular individuals in the world they are about in that they could not exist unless their objects existed as well. In contrast to the class of *de dicto* beliefs, which relate believers to linguistic items, "dicta", perceptual and indexical beliefs are said to be irreducibly *de re*. Beliefs of these last two kinds involve relations between the believer and the relevant things, "res", in her vicinity as part of their identity.

Suppose that, on seeing an animal before me, I say "That animal is a tiger", thereby expressing the demonstrative way of thinking that is made available by my seeing the tiger. What makes my thought a thought about that particular tiger? According to the traditional view, descended from Frege, the way, or *how*, something is represented determines *what*, which object, is represented. The idea is that the subject component of the thought, expressed here by the demonstrative "that animal", imposes a certain condition which an object has to uniquely satisfy in order to be the object of the thought. This view fails, however, because the way we represent an object is often not sufficiently detailed and precise to individuate it fully, that is, to single it out from all other objects. Moreover, sometimes we misrepresent things—what we, at first glance, regarded as a tiger might turn out to be a leopard or a hologram.

Confronted by these sorts of problems, externalists have proposed an alternative picture of the mechanism of reference. In their view, what a singular thought is about is not solely determined by a Fregean sense, a conceptual ingredient of the thought which is sufficient by itself to fix its conditions of satisfaction, but also depends on features of the context in which the thought occurs. Burge contends that a *de re* belief picks out its object in virtue of bearing an appropriate non-conceptual contextual relation to that object.[34] He urges that the content of such beliefs cannot be exhaustively

---

34 Burge 1977

characterized, or completely analysed, in terms of concepts, for the reason that these beliefs contain a crucial contextual element, an element that is independent of any concepts or other types of mental representations in the believer's repertoire. It is the context itself, and not any internal representation of the context, that determines the content of a singular thought. What particular object a person's thought latches onto depends on the circumstances in which the thought occurs and so cannot be derived from anything that happens within the boundaries of her body. It is a consequence of this conception of singular thoughts that they are neither true nor false absolutely, but only relative to their contexts of occurrence. So, in John Perry's instructive portrayal, such thoughts are "essentially indexical".[35] Their content is token-reflexive, contextually variable in the same sort of way in which the content of utterances of sentences containing indexical referring expressions, such as "I am happy today", is.

The Naturalization Project

Since most externalists hold that intentionality is part of the natural order, they are engaged in the naturalization project, the project of reducing mental content, of showing how purely physical states are capable of representing and misrepresenting the world. The central question they have to answer is the question of what it is in virtue of which a mental state has the content it has. What, e.g., makes a particular belief the belief *that p*?

A natural suggestion is to explain content in terms of causal links between mental states and the world. This approach appeals to lawful covariation between mental states and certain external conditions. The core idea is to treat content as a kind of indication or natural meaning in Paul Grice's sense, the kind of meaning that obtains between the number of tree rings and the age in years of the tree, and between red spots and measles. Certain states are indicators of other states. There is a lawful dependency between the indicator and what it indicates. In Dretske's influential information-theoretic account, the point is put by saying that one event carries information about another event in virtue of their being a lawful regularity between them. This notion of carrying information is the basis for the attempt of Dretske and others to clarify how physical states can possess semantic content.

The covariational account, however, is faced with the difficulty of accounting for the possibility of misrepresentation, that is, of error. There can be no misindication. Patently, however, there is misrepresentation. Now and then we form false beliefs. Suppose that beliefs about tigers have regularly been caused by the presence of tigers. But on some occasions beliefs of this type will be caused by things other than tigers, such as leopards, jaguars, holograms of tigers, and so on. Then, if the content of a belief is indeed fixed by what caused it, there is, apparently, no way to misrepresent a leopard as a tiger. A generalization of the problem of misrepresentation is what Fodor calls the "disjunction problem". Simple indicator theories seem to imply that the content of a

---

35 Perry 1993; see also Kaplan 1989

belief of a certain type is a disjunction of the indefinitely many sorts of causes that generate beliefs of that type. If, for example, beliefs of a certain type are caused at different times by tigers, leopards, jaguars and holograms of tigers, then what, on this account, is to prevent beliefs of that type from having the disjunctive content or truth condition: tiger-or-leopard-or jaguar-or-hologram of a tiger?

One might try to meet the disjunction problem by specifying a privileged set of standard or ideal conditions for the formation of beliefs. Then one could identify the content of a belief with its cause under standard conditions, and take false beliefs to be beliefs that arise from other possible causes under non-standard conditions. It has proven extremely difficult, however, to specify such standard or ideal conditions in a non-circular way. In the end, we seem to have no other choice but to construe standard conditions as those in which we form true beliefs. But this gambit is question-begging, for it makes use of the semantic concept of truth.[36]

Dretske tries to meet the disjunction problem in a different way.[37] He equates the content of a belief of a certain type specifically with those conditions that cause tokens of that type during a learning period. In the process of learning to form them, beliefs acquire the semantic content that they have. Misrepresentation is then construed in terms of post-learning-period causes that differ from learning-period causes of beliefs of that type. An evident problem facing this approach is how to demarcate non-circularly the conditions that define learning periods. There seems to be no fixed point at which learning comes to an end. Moreover, a child, while still in the period of learning the concept of a tiger, will, presumably, also be exposed to leopards, jaguars or holograms of tigers, and hence might easily mistake one or the other of these various items for a tiger. So the correlation that learning establishes is between the concept of a tiger and tigers-or-leopards-or-jaguars-or-holograms of a tiger. The original problem returns.

According to teleological theories of content, the disjunction problem can be solved by introducing the notion of biological function or purpose.[38] On this account, beliefs are states with biological functions, and, accordingly, the attempt is made to analyse their contents in terms of those conditions they are biologically designed, or supposed, to covary with. It is these proper functions, designed by natural selection, that fix the content of a belief. The basic idea is that a belief has the content p just in case the belief-forming mechanism which produced it has the function of producing that belief only when it is the case that p. So tiger-beliefs represent tigers, not leopards, because it is their biological function to be held when tigers, but not leopards, are present. It is important that beliefs possess information-carrying functions, even if there are no conditions under which these functions can be carried out appropriately. In such cases our cognitive abilities fail. This is what is supposed to make misrepresentation possible. The beliefs formed in such cases have a false content; they fail to carry the information that, in virtue of the type they are tokens of, it is their natural function to carry. Content and truth fall apart.

---

36   See Cummins 1989; Fodor 1990, 31-49
37   Dretske 1981
38   Millikan 1984; Papineau 1987

Fodor, however, has argued that the teleological account does not really solve the disjunction problem.[39] Suppose that frogs have evolved in environments in which the properties of being a fly and of being a small black dot lawfully covary. Then selectional processes do not show whether it is the biological function of the frog's relevant representation to represent flies or to represent small black dots, or both, or something else which covaries with both. After all, the frog's perceptual system is not malfunctioning when it responds to a small black dot rather than a fly. The determinacy of mental content seems to remain underdetermined by biological function and natural selection.

Fodor himself proposes another solution to the disjunction problem. The essential idea of his "asymmetric dependency" theory is that falsehoods are semantically dependent on truths in a way that truths are not semantically dependent on falsehoods. The symbol "tiger" stands for tigers, and not for tigers-or-leopards, because tigers would still cause tokens of "tiger" even if leopards did not, but leopards would not cause tokens of "tiger" if tigers did not. In general, the causal relations between Ys and tokens of "X" constitute the fact that "X" stands for Ys if for every Z the causal relations between Zs and tokenings of "X" are asymmetrically dependent upon it. Indubitably, the sufficient conditions for mental representations Fodor describes are very plausible. Nevertheless, his reductive account has not obtained universal acceptance. A general worry that is raised is whether the thesis of asymmetric dependency does not implicitly presuppose what it promised to be explaining. Does it more than to give expression to our natural conviction that error is parasitic on truth? Does it really provide us with an explanation of error?[40]

Arguments against Content Externalism

Several renowned philosophers have objected that externalism about content cannot be reconciled with the causal or explanatory role that beliefs, desires and other propositional attitudes are supposed to play in an agent's psychology, that is, with their role in mediating between perception and behaviour.[41] The main argument is based on certain reasonable principles about causation, especially about mental causation. Mental events, so the argument goes, have their effects in virtue of their physical properties, not in virtue of their having content. The fact that the vernacular of folk psychology delineates neural events as takings of attitudes toward contents should not tempt us to think that contents really possess causal potential. In fact, content is causally impotent, screened off by the physical properties of neural events. So the problem that is raised, the problem of mental causation, is to find a causal role for contentful mental states in the aetiology of behaviour. According to externalism, mental content is constituted by relations to items in the non-mental world. And the charge is that these extrinsic relations to things outside of a person's body cannot themselves be involved in the causal

---

39   Fodor 1990
40   See Crane 1995
41   See Dennett 1982, Field 1978, Fodor 1981, Loar 1981, Pylyshyn 1984, Stich 1983

transactions. Causation is local and intrinsic—everywhere. The workings of brains or minds are no exceptions. So the extrinsic relations of a mental state cannot contribute to its causal powers; they cannot be incorporated into the causal mechanisms. Since psychological explanations have to cite causally relevant features of events, and contents are not causes at all, there seems to be no theoretical role left for the notion of content.

Consequently, Jerry Fodor proposes "methodological solipsism" as a "research strategy in cognitive psychology". He regards this as a corollary of his "computational theory of the mind", which is decisively influenced by Alan Turing's systematic conception of computation. Turing's idea of symbol manipulating machines for the first time opened up the possibility of defining any rational computation by simple operations on syntactically structured representations. According to the computational theory of the mind, mental states and processes are computational, that is, have their causal efficacy as a consequence of the formal properties of the representations over which they are defined. Fodor calls this the "formality condition", which psychological theories have to fulfil. Computational processes are formal in the sense that they apply to representations in virtue of their syntax. The crucial point is that being syntactic is a way of not being semantic. Syntactic properties are higher-level physical properties. Therefore, representations can, in virtue of their syntactic structures, causally interact with each other. The syntax of a mental symbol, its shape, is a determinant of its causal role, not its semantic properties. Zenon Pylyshyn takes the same line. He adopts the formality condition, and explicitly says: "cognition is a type of computation".[42]

Stephen Stich is another philosopher of psychology who is convinced that externalist intentional attitude attributions are at odds with the requirements of a comprehensive scientific psychology. He maintains that the propositional attitudes, in virtue of their relational individuation, cannot be incorporated into scientific psychology.[43] According to him, the notion of intentional content is simply out of place in a genuine scientific theory about the mechanisms underlying behaviour. The alternative he offers to folk psychology is the "syntactic theory of the mind". Its basic idea is that causal relations between cognitive states mirror formal relations between syntactic objects. For Stich too, the laws of mental processes appeal to properties of representations that are only syntactic.

Similarly, Daniel Dennett holds that the mind is a "syntactic engine", not a semantic engine.[44] A semantic engine is, he finds, a "mechanistic impossibility". Contents and meanings, according to him, are impotent, merely epiphenomenal with respect to the causal powers of a neural event. They are mere epiphenomena because they cannot influence the mechanistic or syntactic flow of local causation in the nervous system.[45] The predictions we arrive at in the "physical stance" have always hegemony over those of the "intentional stance". Dennett is, or used to be, an antirealist, a sort of instrumentalist, both about the entities purportedly described by folk psychology

---

42 Pylysyn 1984, XIII
43 Stich 1983
44 Dennett 1982
45 Dennett 1991b

and about content. Meanwhile, however, he tends to concede a certain kind of reality to the intentional states.[46]

So, according to methodological solipsism, cognitive psychologists, in trying to explain behaviour, must rely exclusively on what is in the head. Everything that happens outside an agent's head, or beyond the confines of her body, is irrelevant to psychological science. This is so because the causal profile of mental states depends only on their intrinsic syntactical properties. Naturally, causal links with sensory input and behavioural output should not be neglected—although, unfortunately, there is a general inclination in these circles to ignore them. Thus, the main thesis of methodological solipsism is that the systematic role mental states play in the organization of behaviour is just an internal matter of how they interact with each other, with proximal stimuli, and with subsequent bodily movements. Referential relations to states of affairs external to the agent do not figure in psychological explanations. Truth and reference, Fodor proclaims, aren't psychological categories. Psychology is, in this sense, an autonomous discipline.

Methodological solipsism is often buttressed by arguing that molecule for molecule identical counterparts will have the same causal powers and, accordingly, will behave alike regardless of any differences in their environments. The world outside might have been different, or might change, without affecting a person's psychologically relevant properties, the ones that rationalize her behaviour, in the least. As far as psychological explanation is concerned, what counts is how the world is internally represented as being, not how the world really is. It is these modes of representations of things, shaped by a person's subjective point of view, that are constitutive of the causal-explanatory role of mental states. And so, according to the theoretical taxonomy of mental states demanded by methodological solipsism, mental states can be said to be the same or different according as their intra-individual causal role is the same or different.

The Case for Narrow Content

Fodor argues that methodological solipsism does not entail the dismissal of common-sense or folk psychology. In contrast to Dennett's instrumentalism or his more recent "mild realism" and to the various versions of eliminative materialism proposed by Stich, Paul Churchland and Patricia Churchland among others,[47] who all deny that the propositional attitudes are accommodatable within a materialist metaphysics, Fodor insists upon full-fledged intentional realism—the view that there really are intentional attitudes that are causally involved in the production of behaviour. Moreover, he also insists upon scientific intentional realism, that is, he contends, further, that scientific psychology will still have to posit mental states, and will have to contain intentional laws subsuming them, including, in particular, laws that relate beliefs and desires to one another and to actions. For Fodor, mentalistic folk psychology is the breeding-ground of a serious physicalistically acceptable psychology.

---

46   Dennett 1991a
47   See Stich 1983; Churchland, Patricia 1986; Churchland, Paul 1989

Against this background, Fodor shows himself convinced that intentional psychology has to preserve some notion of content—narrow content, constituted by internal aspects of a thought, in contrast to wide or broad content, which comprises the referential relations a thought bears to external entities. Although Oscar's und Twin-Oscar's thoughts have different truth conditions, that is, different wide contents, there is also a sense in which they have the same thoughts, thoughts with the same narrow content. From the inside, their thoughts are semantically indistinguishable from each other. If Oscar were instantaneously and unknowingly transported to Twin Earth, his thoughts would, in a certain sense, remain unchanged. Moreover, it is in virtue of this commonality, this shared content, that they seem to behave in the same way. If, for example, both Oscar and Twin-Oscar are thirsty, then the notion of narrow content can be used to explain and to predict their reaching for a glass of H_2O or XYZ respectively. Would, so it is sometimes asked, an attempt to explain their identical behaviour as resulting from different intentional states not have to invoke a kind of miracle?

It has been objected that for the claim that Oscar's and Twin-Oscar's behaviour is the same to be plausible, "behaviour" must not be construed as intentional behaviour, that is action, since actions themselves are normally individuated externally, in terms of the particular objects or substances they involve.[48] Thus, apparently, "behaviour" can only mean something like bodily motion physically described. However, there seem to be behaviours that are both intentional and relational. And, surely, it is the task of psychology to explain them. So perhaps "behaviour" means, more promising, the type of action both Oscar and Twin-Oscar perform. At any rate, many philosophers of psychology hold that a notion of narrow content, which, by not being answerable to the environment, is able to capture an intentional content common to the protagonists, is just what is needed for the purposes of rationalizing explanation. Without it, so the argument runs, important psychological laws and other generalizations could not be captured.

But what is Narrow Content?

There is an intense debate about how the notion of narrow content should be characterized. Fodor himself takes the narrow content of a mental representations to be a function that maps an external context onto a wide content.[49] So mental representations have the same narrow content just in case they have the same wide content in all the same contexts.[50] It should be noted that, on Fodor's variant of the representational theory of the mind, mental representations are linguistic expressions within a special "language of thought", often referred to as "Mentalese", in which the computations constitutive of cognition occur. Roughly, to believe that p is to stand in a certain relation to a symbol that means that p. One problem facing this approach to narrow

---

48   See Peacocke 1981; Evans 1982
49   See also White 1982
50   Fodor 1990

content is to define the operative concept of context. Furthermore, one would prefer a positive, informative characterization of narrow content, one that goes beyond mere abstraction from wide contents.

A rival and more popular conception of narrow content, "conceptual role semantics" or "functional role semantics", has its roots in the philosophy of language. Sometimes even Wittgenstein, in virtue of his famous dictum that the meaning of an expression is its use, is claimed to be an advocate of this approach.[51] Functional role semantics has salient affinities both with structuralism in linguistics, and with what artificial intelligence researchers call "procedural semantics". Moreover, this approach is strongly influenced by functionalism in the general philosophy of mind, whose chief thesis is that what makes a mental state the type of state it is—what makes a wish a wish—is the functional relations it bears to other mental states, to sensory inputs, and to behavioural outputs. A further important source of inspiration for functional role semantics is the philosophy of Sellars.[52] Sellars argues that there are linguistic rules that license not only moves from linguistic stimuli to linguistic responses—"intralinguistic moves", but also moves from sensory episodes to linguistic responses—"language-entry transitions", and moves from linguistic episodes to extralinguistic bodily movements—"language-departure transitions". And he also accentuates that the network of inferential relations holding between sentences of a public language mirrors the inner network of functional relations holding between thoughts.[53]

The various current versions of functional role semantics in the field hold that narrow content, or narrow meaning, is to be identified with functional or conceptual role. The central idea is that the meaning or content of symbols, public or mental, is fixed by the way they are related to one another.[54] But we must ask exactly what relations are supposed to constitute the content of a representation. The leading accounts attempt to characterize a representation's functional role in terms of its inferential relations, inductive or deductive, to other representation. Of course, most advocates of this approach hope that these inferential connections can be spelled out naturalistically, in terms of purely causal connections. It is worth mentioning that Fodor is among the sharpest critics of this alternative conception of narrow content.[55]

Problems with Narrow Content

The main problem this approach finds itself confronted with is to find a non-arbitrary way of specifying the relevant content-constituting connections, to show exactly how conceptual role engenders the determinate narrow content that mental states allegedly possess. What makes the problem so hard, is the holism of epistemic connections. According to the famous Duhem-Quine thesis, our beliefs about the world, or the sen-

---

51 Harman 1982
52 Sellars 1963
53 Loar 1981
54 Field 1977, 1978; Loar 1981, 1982; Harman 1987; Block 1986; Lycan 1988
55 Fodor 1987; Fodor & LePore 1992; see also Putnam 1988

tences expressing them, are not separately vulnerable to adverse observations, because it is only jointly as a theory that they imply their observable consequences. So once it is admitted that at least some epistemic connections of a belief determine its content, it is hard to avoid concluding that all of them do. Thus, if we follow conceptual role semantics, confirmation holism seems to lead to meaning holism, to the doctrine that the content of a belief is determined by the totality of its epistemic connections. If meaning holism is true, however, then every difference in conceptual role, every theoretical difference, generates a difference in meaning. As a result, it becomes virtually impossible for different people ever to share a belief, or ever to mean the same thing by what they say; it becomes even impossible for two time slices of the same person ever to mean the same thing by what they say. Hence hopes for a serious intentional psychology which, in order to achieve generality, has to quantify over all organisms that are in a specified mental state, would be in vain.

Confronted by such unattractive, destructive consequences, one must try to distinguish a subset of the totality of a belief's epistemic connections as semantically relevant. Any such strategy, however, has to come to grips with the big problem that there seems to be no principled basis for distinguishing between those inferential relations of a representation that constitute its meaning and all of its other inferential relations, that is, for distinguishing between change of belief and change of meaning, between collateral information and determinants of content. In the end, any such attempt to isolate the semantically relevant inferences seems to entail a distinction between analytic statements and synthetic statements, a distinction which, in the face of Quine's devastating criticism of it, most philosophers believe cannot be enforced. Nobody has ever succeeded in clarifying precisely what this dichotomy comes to.

This is not yet the end of the problems for conceptual role semantics. Many philosophers think that it is a desideratum on an adequate semantic theory that meaning determines reference and truth conditions. But, as Twin Earth cases reveal, functional role semantics does not satisfy this desideratum. The thoughts of Oscar and Twin-Oscar have the same functional role, but different reference and truth conditions. To cope with this difficulty, several philosophers have developed two-factor, or dual aspect, semantic theories.[56] The idea underlying this approach is that meaning, or content, consists of two relatively independent components. One component, the internal conceptual role component, gives an account of understanding and cognitive significance. The other component, typically a theory of truth, deals with the properly semantic relations between the representations in the head and their referents in the world. So two-factor theories make a conjunctive claim for each sentence or belief: what its conceptual role is, and what its truth conditions are. Consequently, the words we use in ascribing propositional attitudes have to make a dual contribution: they have to specify both functional role and truth conditions.[57]

---

56 Block 1986; Field 1977; McGinn 1982; Loar 1982; Lycan 1984
57 Loar 1983; Lycan 1988

Not all adherents of conceptual role semantics are convinced that such a supplementation of their theoretical framework is really requisite. Gilbert Harman thinks that a semantic theory with only the one factor of conceptual role will do.[58] In contrast to the prevailing solipsistic versions of conceptual role semantics, he advocates a non-solipsistic version of it, one, according to which his one factor does not stop at the skin, but, by taking into account perception and action, reaches out in the external world of referents. Block aptly portrays Harman's conceptual roles as "long-armed", in opposition to the "short-armed" conceptual roles of two-factor theory he himself favours.[59] Ultimately, Block thinks that the dispute is "merely verbal".

Surely, no reasonable theory can entirely ignore the links between language or mind and the world. If we give up the concepts of reference and truth conditions, we risk giving up linguistic meaning and mental content as well. So it seems that the external relations of language have to be acknowledged anyway. Nonetheless, two-factor theories still have to struggle with the notorious holism problem, since the conceptual role component, the job of which it is to explicate the sense in which there can be considerable differences in meaning in cases where there are no extensional differences, remains a holistic affair. While reference and truth conditions are shared, conceptual role continues to vary idiosyncratically from person to person. Furthermore, these theories are also faced with a new problem, the problem of coordinating the two factors. What is the relationship between narrow content and wide content? In particular, these theories are committed to tell us which conceptual role meshes with which reference and truth conditions in order to constitute the total content of a belief. It is conceded that, in contrast to Fregean sense, conceptual role does not fix reference. So this approach needs, but apparently is not in possession of, some mechanism which keeps narrow content and wide content mutually coherent.

Externalist Replies

The main argument for narrow content turns on the assumption that externalist attitude ascriptions are in conflict with the constraints of a scientifically adequate psychology. Burge has countered that the intentional attitudes, in spite of their relational individuation, are well-suited for a mature psychology. He argues that neither the theoretical commitments nor the language of cognitive psychology, as it is currently practised, conform to the strictures of internalism. Rather, psychological theories—theories of vision, of memory, of belief-formation, of reasoning etc.—deal partly with relations between persons, or animals, and their environments. The point of these theories is to explain in more depth how people see things, how they form beliefs about them, and how they remember them. The relations to the physical world, which are a vital part of these various theories, prompt externalist principles of individuation.

---

58   Harman 1987
59   Block 1986, 636-637

Moreover, Burge attacks the popular charge that externalism violates the allegedly eminently plausible supervenience principle that two persons' propositional attitudes, the determinants of their behaviour, could not be different unless what happens within the boundaries of their bodies were different. In his view, such a principle presupposes rather than establishes that the natural kinds of psychology should be internally individuated. It is not based on facts about the scientific methodology psychologists actually employ, or on a comparison to what counts as a good explanation in mature sciences. Rather, the internalist supervenience principle seems simply to beg the question, since it ignores the important role which the specific nature of a person's or animal's surrounding environment plays in psychological explanations. Attribution of mental states in psychology presuppose a normal background, a background which intentional explanations usually take for granted. The reason why molecule for molecule duplicates may nevertheless differ in respect of their beliefs is that they may have different backgrounds. Burge emphasizes that even the natural sciences employ externalist styles of explanation. Geology and biology, for example, appeal to entities that are not supervenient on their internal material composition. In these sciences too we find a type of individuation of explanatory kinds that makes essential reference to the external relations in which instances of these kinds stand.[60] The crucial point is that whilst causation is local, individuation of natural kinds is not.

Externalism and Self-Knowledge

Many philosophers claim that content externalism poses a serious threat to the traditional doctrine of privileged self-knowledge. There is indeed an asymmetry between our knowledge of our own mental life and our knowledge of the external world—including our knowledge of the mental life of others. We have, as it is often put, first-person authority with regard to the contents of our own mind. We generally know directly or immediately, without recourse to inference from evidence, what we ourselves think, desire, and hope, whereas others can get at the contents of our intentional states only indirectly—by relying on our sayings, actions and the context. The assumption, however, that we have this kind of privileged access to our own minds seems to come into conflict with content externalism's central thesis. If mental content is indeed partly determined by environmental and socio-environmental factors, then knowledge of content should depend on knowledge of these factors. But it cannot reasonably be supposed that these external circumstances will in general be accessible to us—at least not *a priori*, by simple reflection, without empirical investigation. Facts beyond our bodies are as accessible to others as they are to ourselves.

Prima facie, it might seem that we have to give up either the thesis of privileged epistemic access or the tenet of content externalism. This is, no doubt, an unattractive choice for many philosophers. So it is small wonder that several philosophers have a

---

60   See also Davies 1986

sneaking suspicion that the conflict is not genuine. Consequently, they endeavour to espouse a version of compatibilism, according to which content externalism can be reconciled with the doctrine of privileged access to the contents of our own thoughts. In support of such a reconciliation, it might be pointed out that a thinker's second-order beliefs are rationally responsive to her own conscious first-order beliefs, and hence will inherit certain environmental dependencies that shape her first-order beliefs.

Paul Boghossian, Michael McKinsey and others, however, have countered that a compelling case can be made for incompatibilism. So Boghossian is convinced that the conflict between our ordinary conception of authoritative self-knowledge and content externalism is indeed genuine.[61] In particular, he attempts to show that compatibilism implies that we can have a priori knowledge of certain facts about the world, which everyone agrees can be known only on the basis of empirical investigation of the environment. His claim is that compatibilism leads to an absurd overextension in the realm of what can be known from the inside, by purely a priori reasoning.

Other philosophers remain largely unimpressed.[62] Burge has argued that a person's distinctive entitlement to self-knowledge does not conflict with content externalism, because that entitlement is neutral with respect to the issue of whether a person's mental states are individuation-dependent on environmental factors. Since externalism about mental content is a metaphysical thesis whilst privileged access is an epistemological one, there is no reason to be anxious that the two theses will encroach on each other's territory. In his eyes, a person's distinctive entitlement to self-knowledge has its source, first, in the role such knowledge plays in critical reasoning, and second, in the constitutive relation that holds between reflective thoughts and their contents. In cases of self-knowledge, the perspectives of the thinker as subject of the first-order belief and as subject of the second-order belief are unified, and it is this single perspective that forges a rational connection between reflective thoughts and their subject-matter. The entitlement to self-knowledge, Burge contends, is independent of knowledge of the external factors that condition mental content.

Non-Conceptual Content

Beliefs and other propositional attitudes have conceptual contents, contents normally specified by "that ..." clauses. This means, roughly, that, in order to have a belief, a person has to possess the concepts that essentially characterize the content of that belief. Some philosophers assert, however, that not all content is conceptual, that some mental states have non-conceptual contents, contents that do not necessarily involve the possession and application of concepts. One prominent kind of mental state that is claimed to have such non-conceptual content is conscious perceptual experience. While the majority of philosophers still think that the content of perceptual experience is com-

---

61   Boghossian 1989, 1998; McKinsey 1991, 1994
62   See Brueckner 1992; McLaughlin & Tye 1998

pletely conceptual, several philosophers such as Dretske, Evans and Peacocke argue that perceptual content is either completely or at least partly non-conceptual.[63]

The case for non-conceptual perceptual content rests in part on the supposition that the character of experience is phenomenologically much richer and more fine-grained than can be formulated by using concepts at the disposal of an ordinary subject. The non-conceptualists contend that we do not possess, say, as many colour concepts as there are shades of colour we can sensorily discriminate. Our repertoire of colour concepts, expressible by words like "blue" or "red" is too coarse-grained to capture the finest detail, the specific character, of our colour experience.

McDowell, however, points out that to regard the content of experience as non-conceptual is to commit oneself to what Sellars has branded as the "myth of the given"— the view that there is an unconceptualized given element in experience which plays a key role in the acquisition of empirical concepts and which, moreover, functions as an ultimate foundation of our beliefs about the world.[64] McDowell maintains that this picture cannot do justice to the intentionality of perception, and so in the end makes a mystery of the relation between mind and world. In his reply to the phenomenological case for non-conceptual content he concedes that not all qualitative aspects we are presented with in sensory experience can be captured by general concepts. In his countermove, however, he urges that these aspects can be expressed by special demonstrative concepts which are made available by the colour's being given in a particular way in experience. We possess, he insists, conceptual capacities whose content is expressible with the help of demonstrative phrases like "that shade". These conceptual capacities enable the fine-grained sensuous detail which we find in experience, and which is supposed to escape the discriminative powers of our conceptual repertoire, to be embraced within that repertoire. And he is at pains to convince us that what is in play here is a genuine conceptual capacity, one which is not restricted in its exercise to a single occasion in which experience presents one with the respective sample. Rather, he maintains, the capacity is a "recognitional capacity", one which can persist into the future, if only for a short time, and which can, in the absence of the original sample, be exploited in thoughts based on memory.

One can agree with McDowell that we do indeed possess such demonstrative conceptual skills which are made available to the subject by her perceptual state. Such skills permit us to express the content of experience, so that for each sensory feature there is a corresponding concept. But the question arises of whether that fact alone is capable of showing that the content of experience is conceptual through and through, that the very existence of experience depends on the deployment of concepts. One might come to think that the dependency is the other way around, that the demonstrative concepts themselves can only be elucidated in terms of the qualitative content of experience. But if this is so, then the content of experience is independent of the possession of the corresponding concepts.

---

63   Dretske 1981; Evans 1982; Peacocke 1992
64   McDowell 1994; Sellars 1963

A detailed account of non-conceptual perceptual content has been given by Peacocke.[65]According to it, perceptual content is determined by a "scenario", a spatial type under which the region of space around the perceiver must fall if an experience with that representational content is to portray the environment correctly. So the correctness of perceptual content depends upon whether or not the volume of the real world around the perceiver actually instantiates the spatial type that gives the content. The scenario is individuated by a specification of surfaces and features in the environment, and their distances and directions from the perceiver's body as origin. But there is no requirement that the perceiver herself possesses the concepts that enter into characterizing the properties and relations that are constitutive of the scenario. The scenario, lacking any sentence-like structure, is not composed of concepts at all.

Bibliography

Alston, William P., 1989, *Epistemic Justification*, Ithaca: Cornell University Press

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

Block, Ned, 1986, "Advertisement for a Semantics for Psychology", in *Midwest Studies in Philosophy*, Vol. 10, Minneapolis: University of Minnesota Press, 615-78

Boghossian, Paul, 1989, "Content and Self-Knowledge", *Philosophical Topics* 17, 5-26

Boghossian, Paul, 1998, "What the Externalist Can Know A priori", in Wright, Crispin & Macdonald, Cynthia (eds.), *Knowing our own Minds*, Oxford: Clarendon Press, 271-84

BonJour, Laurence, 1985, *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press

Brueckner, Anthony, 1992, "What an Anti-Individualist Knows A Priori, *Analysis* 52, 111-18

Burge, Tyler, 1977, "Belief De Re", *Journal of Philosophy* 74, 338-62

Burge, Tyler, 1979, "Individualism and the Mental", *Midwest Studies in Philosophy*, Vol. 4, Minneapolis: University of Minnesota Press, 73-121

Burge, Tyler, 1986, "Individualism and Psychology", *Philosophical Review* 95, 3-45

Burge, Tyler, 1988, "Individualism and Self-Knowledge", *Journal of Philosophy* 85, 649-63

Chisholm, Roderick, 1966; 2nd ed. 1977; 3rd ed. 1989, *Theory of Knowledge*, Englewood Cliffs, N.J.: Prentice-Hall

Churchland, Patricia, 1986, *Neurophilosophy*, Cambridge, MA: MIT Press

Churchland, Paul, 1989, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press

Crane, Tim, 1995, *The Mechanical Mind*, London: Penguin

Cummins, Robert, 1989, *Meaning and Mental Representation*, Cambridge, MA: MIT Press

Davies, Martin, 1986, "Externality, Psychological Explanation and Narrow Content", *Proceedings of the Aristotelian Society Supplementary Volume* 60, 263-83,

---

65 Peacocke 1992

Dennett, Daniel, 1982, "Beyond Belief", in Woodfield, A., (ed.), *Thought and Object*, Oxford: Clarendon Press, 1-95

Dennett, Daniel, 1987, *The Intentional Stance*, Cambridge, MA: MIT Press

Dennett, Daniel, 1991a, "Real Patterns", *Journal of Philosophy* 88, 27-51

Dennett, Daniel, 1991b, "Granny's Campaigne for Safe Science", in Loewer, B. & Rey, G., (eds.), *Meaning in Mind. Fodor and his Critics*, Cambridge, MA: MIT Press, 87-94

Dretske, Fred, 1971, "Conclusive Reasons", *Australasian Journal of Philosophy* 49, 1-22

Dretske, Fred, 1981, *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press

Evans, Gareth, 1982, *The Varieties of Reference*, Oxford: Clarendon Press

Feldman, Richard, 1985, "Reliability and Justification", *Monist* 68, 159-74

Feldman, Richard & Conee, Earl, 1985, "Evidentialism", *Philosophical Studies*, 48, 15-34

Field, Hartry, 1977, "Logic, Meaning and Conceptual Role", *Journal of Philosophy* 69, 379-408

Field, Hartry, 1978, "Mental Representation", *Erkenntnis* 13, 9-61

Fodor, Jerry, 1981, "Methodological Solipsism considered as a Research Strategy in Cognitive Science", in *Representations*, Hassocks: Harvester Press, 225-253

Fodor, Jerry, 1987, *Psychosemantics*, Cambridge, MA: MIT Press

Fodor, Jerry, 1990, "A Theory of Content and Other Essays", Cambridge, MA: MIT Press

Fodor, Jerry & LePore, Ernest, 1992, (eds.), *Holism: A Shopper's Guide*, Oxford: Basil Blackwell

Foley, Richard, 1987, *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press

Frege, Gottlob, 1919, *Posthumous Writings*, ed. Hermes, H., Kambartel, F., Kaulbach, F., tr. Long, P. & White, R., Oxford: Basil Blackwell

Ginet, Carl, 1975, *Knowledge, Perception, and Memory*, Dordrecht: Reidel

Goldman, Alvin, 1967, "A Causal Theory of Knowing", *Journal of Philosophy* 64, 357-72

Goldman, Alvin, 1976, "Discrimination and Perceptual Knowledge", *Journal of Philosophy* 73, 771-91

Goldman, Alvin, 1979, "What is Justified Belief?", in Pappas, George., (ed.), *Justification and Knowledge: New Studies in Epistemology*, 1-23

Goldman, Alvin, 1986, *Epistemology and Cognition*, Cambridge, MA: Harvard University Press

Harman, Gilbert, 1982, "Conceptual Role Semantics", *Notre Dame Journal of Formal Logic* 23, 242-56

Harman, Gilbert, 1987, "(Non-solipsistic) Conceptual Role Semantics", in LePore, E. (ed.), *New Directions in Semantics*, London: Academic Press, 55-81

Kaplan, David, 1989, "Demonstratives", in Almog, J., Wettstein, H. & Perry, J., (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 481-563

Kripke, Saul, 1980, *Naming and Necessity*, Oxford: Basil Blackwell

Loar, Brian, 1981, *Mind and Meaning*, Cambridge: Cambridge University Press

Loar, Brian, 1982, "Conceptual Role and Truth-Conditions", *Notre Dame Journal of Formal Logic* 23, 272-83

Loar, Brian, 1983, "Must Beliefs be Sentences?", in Asquith, P.D., & Nickles, T., (eds.), *PSA 1983*, Vol. 2, East Lansing, MI: *Philosophy of Science Association*, 627-43

Lycan, William, 1984, *Logical Form in Natural Language*, Cambridge, MA: MIT Press

Lycan, William, 1988, *Judgement and Justification*, Cambridge: Cambridge University Press

McDowell, John, 1994, *Mind and World*, Cambridge, MA: Harvard University Press

McGinn, Colin, 1977, "Charity, Interpretation and Belief", *Journal of Philosophy* 74, 521-35

McGinn, Colin, 1982, "The Structure of Content", in Woodfield, A., (ed.), *Thought and Object*, Oxford: Clarendon, 207-58

McKinsey, Michael, 1991, "Anti-Individualism and Privileged Access", *Analysis* 51, 9-16

McKinsey, Michael, 1994, "Accepting the Consequences of Anti-Individualism", *Analysis* 54, 124-28

McLaughlin, Brian & Tye, Michael, 1998, "Externalism, Twin Earth, and Self-Knowledge", in Wright, Crispin & Macdonald, Cynthia (eds.), *Knowing our own Minds*, Oxford: Clarendon Press, 285-320

Millikan, Ruth, 1984, *Language, Thought and Other Biological Categories*, Cambridge, MA: MIT Press

Nozick, Robert, 1981, *Philosophical Explanations*, Oxford: Oxford University Press

Papineau, David, 1987, *Reality and Representation*, Oxford: Basil Blackwell

Peacocke, Christopher, 1981, "Demonstrative Thought and Psychological Explanation", *Synthese* 49, 187-217

Peacocke, Christopher, 1992, *A Study of Concepts*, Cambridge, MA: MIT Press

Perry, John, 1993, "The Problem of the Essential Indexical", *Nous* 13, 3-21

Plantinga, Alvin, 1993a, *Warrant: The Current Debate*, Oxford: Oxford University Press

Plantinga, Alvin, 1993b, *Warrant and Proper Function*, Oxford: Oxford University Press

Pollock, John, 1986, *Contemporary Theories of Knowledge*, Totowa, N.J., Rowman & Littlefiedl

Putnam, Hilary, 1975, "The Meaning of 'Meaning'", in *Mind. Language and Reality*, Cambridge: Cambridge University Press, 215-271

Putnam, Hilary, 1988, *Representation and Reality*, Cambridge, MA: MIT Press

Putnam, Hilary, 1992, "Replies", *Philosophical Topics* 20,347-408

Pylyshyn, Zenon, 1984, *Computation and Cognition*, Cambridge, MA: MIT Press

Ramsey, Frank, 1931, "Knowledge" in Ramsey, *The Foundations of Mathematics and Other Essays*, ed. Braithwaite, R.B. New York: Harcourt Brace

Schiffer, Stephen, 1977, "Naming and Knowing", in *Midwest Studies in Philosophy*, Vol. 2, Minneapolis: University of Minnesota Press, 28-41

Searle, John, 1980, "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3, 417-24

Searle, John, 1983, *Intentionality*, Cambridge: Cambridge University Press

Sellars, Wilfrid, 1963, "Empiricism and the Philosophy of Mind", in *Science, Perception and Reality*, London: Routledge & Kegan Paul, 127-96

Sosa, Ernest, 1991, *Knowledge in Perspective*, Cambridge: Cambridge University Press

Stich, Stephen, 1983, *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press

Swain, Marshall, 1981, *Reasons and Knowledge*, Ithaca: Cornell University Press

Talbott, William, 1990, *The Reliability of the Cognitive Mechanism*, New York: Garland Publishing

White, Stephen, 1982, "Partial Character and the Language of Thought", *Pacific Philosophical Quarterly* 63, 347-65

I
In Defence of Epistemic Externalism

The "Challenge" of Externalism

William P. Alston

I

The title of this volume reflects a widespread assumption among philosophers that externalism, whether in epistemology or in philosophy of mind, is something radically new on the scene. This new entry might be viewed as either dangerous or hopeful, depending on the philosophical orientation. In this essay I will be concerned exclusively with *epistemological* externalism. I will argue that the sense of a radical departure from "the epistemological tradition" stems from (1) an oversimplified picture of that tradition and (2) a failure to realize the complexity of contemporary externalism. I will not go so far as to suggest that externalism poses no challenge. It is far from toothless. But the challenge is not exactly what it is often credited with being.

I begin with a statement of the challenge from an internalist opponent. In BonJour 1985 he writes.

> When viewed from the general standpoint of the Western epistemological tradition, externalism represents a quite substantial departure. It seems safe to say that until very recent times, no serious philosopher of knowledge would have dreamed of suggesting that a person's beliefs might be epistemically justified merely in virtue of facts or relations that are external to his subjective conception. Descartes, for example, would surely have been quite unimpressed by the suggestion that his problematic beliefs about the external world were justified if only they were in fact reliably caused, whether or not he had any reason for thinking this to be so. Clearly his conception, and that of generations of philosophers who followed, was that such a relation could play a justificatory role only if the believer himself possessed adequate reasons for thinking that the relation obtained. Thus the suggestion embodied in externalism would have been regarded as simply irrelevant to the main epistemological issue, so much so that the philosopher who suggested it would have been taken either to be hopelessly confused or to be simply changing the subject (...this *may* be what some externalists intend to be doing). (p. 37)

And here is a statement by the dean of American epistemology in the second half of the 20th century, Roderick Chisholm.

> We must be on guard...in interpreting contemporary literature that professes to be about "internalism" or "externalism". Some of those authors who profess to view knowledge and epistemic justification "externally" are not concerned with traditional theory of knowledge. That is to say, they are not concerned with the Socratic questions, "What can I know?", "How can I be sure that my beliefs are justified?", and "How can I improve my present stock of beliefs?". Indeed many such philosophers are not concerned with the analysis of any ordinary concept of knowledge or of epistemic justification. Therefore their enterprise, whatever it may be, is not that of traditional theory of knowledge. (Chisholm 1989, pp. 75-76)

To evaluate these claims we must consider what "the Western epistemological tradition" or "traditional theory of knowledge" has to say, or is assuming, about matters in dispute between contemporary internalism and externalism. But first we must settle on a way to construe these current positions.

II

The terms 'internalism' and 'externalism' are used in epistemology, as elsewhere, in various ways. Thinking of them as applied to theories of epistemic *justification*, as is commonly done, the contrast that is most prominent in the literature has to do with whether it is a necessary condition of something's being a "justifier" of a belief, i.e., something that contributes to its justification, that the subject have some kind of high grade cognitive access to it. Internalism gives a positive answer, externalism a negative answer to this question. (A second contrast will be introduced in due course.) With respect to this first contrast, 'internalism' is the term that, as J. L. Austin put it in the old male chauvinist days, "wears the trousers". Internalism puts certain restrictions on what can contribute to justification, and, on this point, externalism is simply the denial of those restrictions.[1]

Before moving on to the second contrast the first one needs to be sharpened up. First, there is a distinction between mere epistemic *access* and the successful exploitation thereof. That is, we might think of internalism as holding that epistemizers are restricted to what the subject has some high grade of access to, or what the subject actually has knowledge or justified belief concerning. In Alston 1986 the former is called "access internalism" and the latter "perspectival internalism". BonJour's phrase, "one's subjective conception" suggests the latter. (Only "suggests", because one might have a conception of a fact without knowing or being justified in believing that any such fact obtains. But I take it from what BonJour has written elsewhere that he really intended the epistemically stronger version.) But the "access" version is more prominent in the literature. It would generally be thought that reasons for a belief would pass the internalist test provided the subject could fairly easily become aware of them, even if the subject had never actually incorporated them into his belief system. Moreover, the access version is not so obviously saddled with an infinite regress as the perspectival version. (If one can be justified in believing that *p* only if one is justified in believing that conditions for being so justified are satisfied, then that belief in turn is subject to a similar requirement, and so on ad infinitum. This implies that one cannot be justified in any belief without possessing an infinite hierarchy of justified beliefs.) Hence I will conduct this discussion in terms of access internalism.

That still leaves at least two alternatives dangling. First, how strong does the access have to be? If it is merely that the subject must be able, somehow, to ascertain that the justifiers obtain, it is doubtful that any epistemizing conditions favored by externalists,

---

1   Later we will see that something more positive is involved.

e.g., reliability of belief formation, would be excluded. Much stronger versions of the access requirement have been proffered. For example, in Ginet 1975, Ch. III, it is required that any justifier be recognizable just by clear headedly asking oneself whether it does obtain. Rather than spend the entire essay on this issue, I will think in terms of a vague requirement of "relatively direct ascertainability". The other alternative is between the existence of the alleged justifier and its justificatory efficacy. One might well have strong access to the former and not the latter. Presumably the existence of a sensory experience is directly ascertainable by its possessor, but knowledge of its justificatory function vis-a-vis a perceptual belief is rather harder to come by. Much of the internalist literature suffers from ignoring this distinction. But since internalist concerns would not be fully accommodated by the right kind of access to the existence of the putative justifier while being unable to determine whether it deserves that title, I will think in terms of strong access to both being required by internalism.

As previously noted, we may think of externalism, in the first instance, as the denial of this restriction. Justifiers need not be fairly directly ascertainable by the subject, either their existence or their justificatory efficacy. This does *not* imply that externalism holds that justifiers always or usually fail to satisfy the internalist restriction. It only denies that it is *necessary* for their justifying status that they do so. I mention this at the risk of overkill because misapprehension on this point is not uncommon. And this is particularly unfortunate, since externalists typically recognize justifiers that do satisfy an internalist restriction on access, e.g., sensory experiences and conscious reasons.

III

Now I turn to the question of whether the "Western epistemological tradition" is uniformly internalist on the access issue. It would be a miracle if there were complete unanimity on this point, as well as on any other point, among all philosophers prior to the 20th century who have been concerned with what today we identify as epistemological issues. In order to keep the discussion within manageable bounds, I will restrict myself to the 17th and 18th centuries, which have largely set the tone for subsequent epistemology through the 20th century.

In attempting to consider whether the major epistemologists of this period were internalists in anything like the way BonJour and Chisholm are construing internalism, the first roadblock we run into is this. Both philosophers are typical of contemporary Anglo-American epistemologists in supposing that the "Western epistemological tradition" has at all times shared the twentieth century preoccupation with "justification" as a prime epistemic desideratum and as a necessary condition for a true belief's counting as knowledge. But there are a number of reasons why Descartes and Locke, for example, along with most other pre-twentieth century philosophers, are innocent of any such tendencies. The most important point is that so far from thinking that "justification" is required for turning a true belief into knowledge, they would deny that anything could do that trick. They took knowledge to be a psychological state of a radically different sort from belief (judgment, opinion......). Knowledge was usually thought of as some kind of *immediate awareness* of facts, rather than as a *judgment* that a certain

fact obtains. One place in Descartes where this is particularly clear is Meditation IV, where he repeatedly refers to knowledge as a function of the understanding and to judgment as an exercise of the will. The main point he makes there is that it is incumbent on us, when judging, to follow the lead of the "clear and distinct perceptions" of the understanding in what we judge to be the case. Locke is much more explicit about meta-epistemology than Descartes. "*Knowledge* then seems to me to be nothing but *the perception of the connexion and agreement, or disagreement and repugnancy of any of our Ideas*. In this alone it consists." (Essay, IV, 1, ii).[2] In chapter 14 of Bk. 4 he is at great pains to distinguish knowledge from judgment. "The Faculty, which God has given Man to supply the want of clear and certain Knowledge in Cases where that cannot be had, is Judgment: whereby the Mind takes its Ideas to agree, or disagree; or which is the same, any proposition to be true, or false, without perceiving a demonstrative Evidence in the Proofs." (IV, 14, iii). Hume in the chapter on Knowledge in the *Treatise* follows Locke in this. "It appears, therefore, that of these seven philosophical relations, there remain only four, which depending solely upon ideas, can be the objects of knowledge and certainty". (I, iii, 1) By "depending solely upon ideas" Hume means that those relations necessarily follow from the nature of the relata, and hence can be ascertained with certainty just by being aware of the ideas so related. What cannot be known in this way is, as in Locke, relegated to mere belief.

By the time we reach the late 18th century we find both Kant and Reid treating knowledge as a judgment that satisfies certain conditions. Reid does us the favor of explicitly criticizing Locke's distinction and insisting that "there can be no knowledge without judgment, though there may be judgment without that certainty which we commonly call knowledge." (VI, 1.)

But even though the treatment of knowledge as distinct from all judgment is not omnipresent in these centuries, that is not of much aid and comfort to those who would foist our 20th century concern with "epistemic justification" on their early modern predecessors. For we will not find them giving prominence to any concept very close to our concept of epistemic justification, in any of its many versions. I do not suggest that any important philosopher has been unaware that some of our beliefs (judgments) are better grounded, founded, supported, than others; and various terms are used for this. Locke, for example, speaks of degrees of probability. But even so, aside from Locke, we do not find prior to Kant and Reid a great deal of interest in cognitive achievements that fall short of knowledge. Such attention as they give to determinants of "positive epistemic status", to use as neutral a term as possible, is mostly focused on what makes for knowledge.

How, then, can we evaluate BonJour's claim without distorting either that claim or our 17th and 18th century forebears? So far as I can see, the closest we can come is

---

2   I am not concerned here with Locke's restriction of the object of knowledge to one's own ideas, something to which, in any case, he does not consistently adhere. My concern here is with the way in which he restricts knowledge to an immediate awareness of facts, distinguishing it from judgment (belief) in which some proposition is taken to be true without our being directly aware of a fact that makes it true.

this. Let's coin the term 'epistemizer' for anything that engenders a positive epistemic status for any propositional attitude, whether knowledge or belief. We can then ask whether the 17th and 18th century epistemologists subscribed, at least in practice, to access internalism with respect to epistemizers. But even here we must tread cautiously. Where knowledge is concerned everyone, no matter how internalist, will recognize that knowledge that p requires that it is true that p. And it is not generally true that people have a relatively direct access to the truth of all the true proposition we can envisage. If that were the case, we would all know a great deal more than we do. Hence it would hardly be reasonable to use a direct access requirement for everything that contributes to knowledge (including truth) as a litmus test for access internalism.

But, of course, since it is (almost) universally agreed that knowledge requires more than true belief (or true whatever might or might not be knowledge), we could get around this difficulty by restricting our litmus test to whatever contributes to the knowledge status other than truth. But there are problems with that also. Contemporary internalists who think that justification and truth are required for a belief's counting as knowledge have been sufficiently impressed by Gettier to recognize that they are not sufficient. In addition, something must be there that will obviate Gettier problems. And, again, internalists accept, for good reasons, that we lack direct access to the satisfaction of these anti-Gettier conditions, e.g., the absence of any fact that when added to the justifying conditions would result in the belief's not being sufficiently justified. Hence we can't simply say: an epistemologist is an access internalist about knowledge provided he holds that anything that contributes to knowledge status other than truth is something to which we have relatively direct access. And so we still lack an acceptable general formulation.

Recall what is giving us this difficulty. The aspect of knowledge to which contemporary internalists require relatively direct access is "justification". But most 17th and 18th century epistemologists, especially those who take knowledge and belief to be disjoint, do not suppose that anything very close to 20th century justification to be a necessary condition of knowledge. Therefore we have fishing in vain for something connected to knowledge with respect to which they might or might not conform to the image those like BonJour and Chisholm have of them as access internalists. But all is not lost. After all, these ancestors of ours do take there to be something that is required for knowledge other than truth. What prevented us from saying that they fit BonJour's image (as far as knowledge as concerned) *iff* they took *what turns a true whatever into knowledge* to be directly accessible, was that for our contemporaries this also includes some condition to block Gettier problems. To be sure, we might rule that as irrelevant to the issues, since the 17th and 18th century epistemologists had the misfortune of disappearing from the scene before Gettier's epoch making 1963 article and hence had no chance to take any position on cognitive accessibility to Gettier-problem-blocking conditions. But we can secure the same result in a less high-handed way. Let's say that they are accessibility internalists with respect to knowledge *iff* they suppose that the subject has relatively direct access to anything constitutive of knowledge other than truth and what it takes to block Gettier problems. If they are innocent of any concern with the latter, we can still consider whether they are internalists with respect to other knowledge-making features.

I'm afraid that is the best I can do to make BonJour's claim into something that can be genuinely investigated. To sum up, the claim can be construed as follows.

> BJ: The epistemological tradition takes it that any contributor to positive epistemic status for beliefs (judgments) and any contributor to a knowledge status, other than truth and what it takes to block Gettier problems, is relatively directly accessible to the subject.

Remember that the discussion is conducted on the fiction that "the epistemological tradition" is confined to the 17th and 18th centuries.

Now that we have a question, what is the answer? Without trying to go into the matter in the way that would be required for a serious historical investigation, I would suppose that it is positive. For Descartes what makes for knowledge is "clear and distinct perception", and there is every reason to suppose that he took it that having such a perception carries with it a (perhaps infallible) capacity to spot it. When in Meditation III he comes to raise doubts about whether such perception yields knowledge, they are not based on any uncertainty as to when one has such perceptions, but rather on whether we are so constituted that they are a reliable guide to truth. Even one as determined to surmount every possible sceptical doubt as Descartes never worries about our ability to recognize clear and distinct perception. Locke, in parallel fashion, never suggests that one might fail to register a "perception of the agreement or disagreement of ideas", or might suppose one has such a perception when one doesn't. These are facts of a sort that are wholly open to our awareness. Nothing can get between me and facts like this that could make me miss them or hallucinate them. Something of the same sort could be said for Hume (where knowledge is concerned) and for other advocates of an intuitive conception of knowledge.

The matter is less clear with respect to what makes for positive epistemic status of belief or judgment, mostly, I think, because this receives much less attention in the period. But my sense is that there too BonJour's thesis holds up. I believe that a close reading of, e.g., Locke and Reid on this point would again support BJ. For Locke (Essay, IV, xvi, 4), the "grounds of probability" are "conformity of any thing with our own Knowledge, Observation, and Experience" and "the Testimony of Others, vouching their observation and Experience". Though the latter, involving sense perception, is obviously less directly accessible than the former, I believe that Locke would take it to be sufficiently directly and unproblematically available to satisfy BonJour's supposition. And Reid, who took the forms of "evidence" that ground our judgments to be sense, memory, consciousness, or credit in testimony, seems for all the world to suppose (though I am unaware of his ever explicitly discussing the matter) that these "evidences" are unproblematically and relatively directly accessible to the subject of the judgment in question.

I have already acknowledged the slapdash character of this survey. But since I am primarily concerned in this essay to contest the claim that "the epistemological tradition" is uniformly internalist, my excuse for superficiality is that since here I agree with BJ my position is one of *acknowledging* a point made by my opponent. Hence I am under no obligation of supporting the position as solidly as is required when I dissent from my interlocutor.

IV

So if the denial of access internalism were all there is to externalism, I would have to
agree with BonJour and others that externalism makes a sharp break with the epistemo-
logical tradition. But externalists are generally as deeply committed to a more positive
feature—the insistence that epistemizers be *truth-conducive*. (I will use 'TC' to abbrevi-
ate 'truth-conducivity'.). The view that epistemizers are essentially TC is the view that
nothing can function as an epistemizer unless it contributes to making the propositional
attitude in question at least likely to be true. Thus reliabilism, the most common ex-
ternalist position, holds that a belief is justified (alternatively that it is has what it takes
for knowledge other than truth and something to deal with Gettier problems), only
if it has been formed in a *reliable* way, a way that can be relied on to produce mostly
true beliefs in a wide spread of situations of the sort we ordinarily encounter. The same
holds for such forms of externalism as Nozick's "tracking" requirement for knowledge,
and Plantinga's view of "warrant" as belief acquisition by the *proper functioning* of cog-
nitive faculties, though it would take more space than I have to show this.

Moreover, TC is not just externally conjoined with the denial of strong access. They
are intimately connected. It is because externalists hold that the point of having justi-
fied rather than unjustified beliefs is that we want our beliefs to at least be likely to be
true, that they reject strong access requirements. They are convinced that much of what
renders beliefs likely to be true is not directly accessible to reflection, introspection, or
the like. Thus they take there to be a choice between a direct access requirement and
TC, and they choose the latter.

To be sure, the TC aspect of externalism does not present such a clear cut opposi-
tion to internalism as the denial of an access requirement. Some internalists, including
BonJour, are equally committed to TC. At another place in BonJour 1985 he writes:

> If epistemic justification were not conducive to truth in this way, if finding epistemically jus-
> tified beliefs did not substantially increase the likelihood of finding true ones, then epistemic
> justification would be irrelevant to our main cognitive goal and of dubious worth. (p. 8)

But other internalists, including Chisholm, Ginet, and Foley are equally insistent that
there is no conceptual connection between justification (Foley says 'rationality') and
the likelihood of truth.

> According to this traditional conception of "internal" epistemic justification, there is no *logi-*
> *cal* connection between epistemic justification and truth. (Chisholm 1989, p. 76)[3]
> Epistemic rationality is distinguished from other kinds of rationality by its truth-directed
> goal. The goal is for one now to believe those propositions that are true and now not to be-
> lieve those propositions that are false.
> However, to say that the goal that helps distinguish epistemic rationality from other kinds
> of rationality is a truth-directed goal is not to say that truth is a prerequisite of epistemic
> rationality. In particular, it is not to say that it is impossible for what is epistemically rational
> to be false; and likewise it is not even to say that it is impossible for most of what is epis-
> temically rational to be false. (Foley 1987, p. 155)

---

3   Chisholm is not just making the uncontroversial point that justification does not entail truth, but the
    stronger claim that it does not entail even the likelihood of truth.

Because externalism includes this positive feature of truth-conducivity, as well as the negative rejection of internalist access requirements, in deciding whether externalism is a radical departure from the epistemological tradition, we must take both aspects into account. We have already seen that a good case can be made for its novelty with respect to the second, negative aspect. But what about the positive aspect?

Here the situation is radically different. Go back to Descartes' Meditation III, where, as I indicated in the last section, Descartes' reason for his initial doubts as to whether clear and distinct perceptions provide knowledge is a doubt as to whether they are a reliable source of truth. It is worth while examining the passage in detail.

First, Descartes goes over the things he is convinced that he knows with certainty and notes that in every case there is "a clear and distinct perception of what I am asserting". That leads him to "lay it down as a general rule that whatever I perceive very clearly and distinctly is true". But before he can allow himself to rest in this reassuring confidence a doubt insinuates itself. "...this would not be enough to make me certain of the truth of the matter if it could ever turn out that something which I perceived with such clarity and distinctness was false". But is there any chance that this should happen? Here the idea of a deceiving deity suggest itself.

> Did I not see at least these things clearly enough to affirm their truth? Indeed, the only reason for my later judgment that they were open to doubt was that it occurred to me that perhaps some God could have given me a nature such that I was deceived even in those matters which seemed most evident. And whenever my preconceived belief in the supreme power of God comes to mind, I cannot but admit that it would be easy for him, if he so desired, to bring it about that I go wrong even in those matters which I think I see very clearly...And since I have no cause to think that there is a deceiving God, and I do not even know for sure yet whether there is a God at all, any reason for doubt which depends simply on this supposition is a very slight and, so to speak, metaphysical one. But in order to remove even this slight reason for doubt, as soon as the opportunity arises I must examine whether there is a God, and, if there is, whether he can be a deceiver. For if I do not know this, it seems that I can never be quite certain about anything else.[4]

Descartes is then led into the famous "Cartesian circle", in which he seeks to remove the above doubt concerning the reliability of clear and distinct perceptions by showing that there is an omnipotent and perfectly good God and that we are endowed with our cognitive faculties by Him. In so doing he relies on clear and distinct perceptions to provide the premises of the argument, thus assuming their reliability in order to show that they are reliable. I have no concern here with that venerable problem. It suffices for present purposes to point out that Descartes will not recognize the genuineness of any putative source of knowledge unless he can show that it is maximally truth-conducive, i.e., infallible. This is to demand the highest degree of truth-conducivity of anything that will contribute to a particular cognition's counting as knowledge. And note that Descartes not only requires that clear and distinct perception *be* infallible, but that he *know* that they are, thus combining the externalist requirement of truth-conducivity and the internalist requirement of higher level knowledge of this.

---

4   The text is that in Cottingham, Stoothoff, and Murdoch (trans.) 1984, Vol. II, p. 25.

On the present point Locke is less explicit than Descartes, the reverse order having been noted on the relation of knowledge and belief. But I take it to be sufficiently clear that Locke holds that what makes a true cognition into a case of knowledge is such as to *guarantee* that truth, again the highest degree of truth conducivity. To wit.

> When we possess our selves with the *utmost security* (emphasis added) of the Demonstration, that *the three Angles of a Triangle are equal to two right ones*, What do we more but perceive that Equality to two right ones, does necessarily agree to, and is inseparable from the three Angles of a Triangle?" (IV, i, 2)
> By this the Mind clearly and *infallibly* (emphasis added) perceives each *Idea* to to agree with itself and to be what it is, and all distinct *Ideas* to disagree, i.e., the one not to be the other. (IV, i, 4)

And speaking of knowledge of modes,

> So that we cannot but be *infallibly* (emphasis added) certain, that all the knowledge we attain concerning these *Ideas* is real, and reaches Things themselves. Because in all our Thoughts, Reasonings, and Discourses of this kind, we intend Things no farther, than as they are conformable to our Ideas. So that in these, we cannot miss of a certain undoubted reality. (IV, iv, 5)
> Wher-ever we perceive the Agreement or Disagreement of any of our *Ideas* there is certain Knowledge: and where-ever we are sure those *Ideas* agree with the reality of Things, there is certain real Knowledge. (IV, iv, 18)

It is clear from the larger context that 'certainty' in these passages denotes not feeling sure or confident, or entertaining no doubt, but rather what we might term *epistemic certainty*, being in such a position that it is impossible that one should be mistaken.

I could also cite the same position, that what makes for knowledge endows the cognition with certainty in the above sense, for Hume, Kant, and others.

Things are less clear cut, again, with views on the truth-conducivity of determinants of positive epistemic status for judgments in this period. But I believe that it is sufficiently clear that the leading figures were construing these determinants as exhibiting a built-in truth conducivity. Begin with Locke. His term of choice for what I am calling a determinant of "positive epistemic status" for judgments is *probability*. Though Locke does not say as much as we would like about how he thinks of probability, he says enough to indicate that he thinks a strong ground of probability to be a significantly reliable indication of truth. Here is the relevant portion of the first section of Ch. XV of Bk. IV, entitled "Of Probability".

> As Demonstration is the shewing the Agreement, or Disagreement of two *Ideas*, by the intervention of one or more Proofs, which have a constant immutable and visible connection one with another; so *Probability* is nothing but the appearance of such an Agreement, or Disagreement, by the intervention of Proofs, whose connexion is not constant and immutable, or at least is not perceived to be so, but is, or appears for the most part to be so, and is enough to induce the Mind to judge the Proposition to be true, or false, rather than the contrary. (IV, xv, 1)

This is not lucidly expressed and requires some imaginative exegesis. The syntax would seem to indicate plainly what is spoken of as having (or not having) "an immutable and visible connection one with another" is the proofs. But the passage makes sense only if this phrase applies to the ideas in question, and so I will read it. Where we "perceive" an agreement or disagreement of ideas, there can be no question that such a relation

obtains. But where the "proofs" show only a probability of such a connection, it "is, or appears for the most part to be so", and so the mind is induced to judge the proposition asserting such a connection to be true. Thus far Locke has not clearly committed himself to an objective truth-conducivity of the probability established by the proofs. The disjunction of 'is' with 'appears' blocks any such implication. And the reference to what "induces the mind" suggests persuasive efficacy, rather than a reliable indication of truth. But in an example later in the section Locke is more explicit as to a conceptual connection between probability and likelihood of truth. He has just described a person who had demonstrated the theorem and continues with the contrasting case of one who makes a probable judgment.

> But another Man who never took the pains to observe the Demonstration, hearing a Mathematician, a Man of credit. affirm the three Angles of a Triangle, to be equal to two right ones *assents* to it; i.e., receives it for true. In which case, the foundation of his Assent is the Probability of the thing, the Proof being such, as for the most part carries Truth with it: The Man on whose Testimony he receives it, not being wont to affirm any thing contrary to, or besides his Knowledge, especially in matters of this kind. (IV, xv, 1)

This says as clearly as possible that a (sufficiently) probable judgment is one whose grounds constitute a reliable indication of truth. For here the proof is "such as for the most part carries Truth with it". And it does so because of the reliability of the testifier "in matters of this kind". Not only do we find Locke taking the determinants of positive epistemic status for judgments to be truth-conducive. His way of committing himself to this would seem to qualify him as a reliability theorist of judgment in a way that would be congenial to contemporary externalists. And at the beginning of section 3 of the chapter he nails this down by saying "Probability is likeliness to be true..."

The other witness I will call for the prosecution is Thomas Reid. The most relevant material is found in his discussion of what he calls "first principles" in the *Essays on the Intellectual Powers of Man,* or at least such of these as can be construed as principles for epistemically evaluating judgments. They all have to do with the reliability of various grounds of judgments (sorts of "evidence", to use Reid's favorite term). Some of the principles take the form of affirming the real existence of what a given cognitive faculty seems to reveal.

> I shall take it for granted, that I *think*, that I *remember*, that I *reason*, and, in general, that I really perform all those operations of mind of which I am conscious. (I, 2, p. 34)[5]
> ...those things did really happen which I distinctly remember. (VI, 5, p. 622)
> ...those things do really exist which we distinctly perceive by our senses, and are what we perceive them to be. (VI, 5, p. 625)
> ...I shall also take for granted such facts as are attested to the conviction of all sober and reasonable men, either by our senses, by memory, or by human testimony. (I, 2, p. 40)

Sometimes the principle is in terms of the reliability of a source of knowledge.

> As by consciousness we know certainly the existence of our present thought and passions, so we know the past by remembrance. (I, 2, p. 34)

---

5   Page references are to the edition published by the MIT Press (1969).