Internalism and Epistemology

The Architecture of Reason

Timothy McGrew and Lydia McGrew

Routledge Series in Contemporary Philosophy

Internalism and Epistemology

Internalism and Epistemology is a powerful articulation and defense of a classical answer to an enduring question: What is the nature of rational belief? In opposition to prevailing philosophical fashion, the book argues that epistemic externalism leads, not just to skepticism, but to epistemic nihilism – the denial of the very possibility of justification. And it defends a subtle and sophisticated internalism against criticisms that have widely but mistakenly been thought to be decisive.

Beginning with an internalist response to the Gettier problem, the authors deal with the problem of the connection to truth, stressing the distinction between success and rationality as critical to its resolution. They develop a metaregress argument against externalism that has devastating consequences for any view according to which epistemic principles are contingent. The same argument does not, they argue, affect the version of internalism they espouse, since its epistemic principles are analytic and knowable *a priori*. The final chapter addresses the problem of induction and shows that its solution turns critically on the distinction between success and rationality – the very distinction that lies at the heart of the dispute between internalists and externalists.

Provocative, probing, and deliberately unfashionable, *Internalism and Epistemology* is a ringing defense of internalism that will interest specialists and students alike. It is essential reading for anyone who suspects that rumors of the death of traditional epistemology have been greatly exaggerated.

Timothy McGrew is Professor and Chairman of the Department of Philosophy at Western Michigan University and the author of *The Foundations of Knowledge* (1995). His articles have appeared in numerous journals including *Analysis*, *Mind*, *The Monist*, and *British Journal for the Philosophy of Science*.

Lydia McGrew lives in southwestern Michigan where she educates her three children and continues her scholarly work in epistemology and probability theory. She has published philosophy articles in journals including *Australasian Journal of Philosophy*, *Philosophia Christi*, and *Mind* (with Timothy McGrew).

Internalism and Epistemology

The Architecture of Reason

Timothy McGrew and Lydia McGrew

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2007
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been applied for

ISBN 0–203–96830–1 Master e-book ISBN

ISBN10: 0-415-77067-X (Print Edition)
ISBN13: 978-0-415-77067-5

Toute notre dignité consiste donc en la pensée. . . . Travaillons donc à bien penser.

Pascal

Contents

Introduction

Though the word 'philosophy' means the love of wisdom, what interests philosophers most is reason. And small wonder; for reason, considered in its own right, is both the primary tool and the special province of philosophy. Philosophers reason about various matters of fact – the existence of God, the external world, and other minds, for example – but they also reason about pieces of reasoning. And at the deepest levels they reason about reason itself, its scope and limits, its various structures, its foundations, and its relation to the concepts of truth, justification, and knowledge.

It matters a great deal, then, when a contemporary philosophical dispute turns on a choice between profoundly different conceptions of the nature of reason and of the relation of reason to knowledge. A decision on a matter of such moment will have a decisive bearing on one's conception of philosophy itself. Such is the dispute in contemporary epistemology between the partisans of internalism and externalism.

As often happens in philosophy, the key terms are defined in multiple ways in the literature. But beneath the multiplicity of definitions lies a fundamental difference in stance, a difference in what each side values. Advocates of internalism stress the connection between rational belief and knowledge. They take the traditional skeptical challenges seriously, and they believe that a major part of the epistemologist's task is to grapple with those challenges and, where possible, to overcome them. Externalists, by contrast, are generally uninterested in skeptical questions, often because they consider the game to be rigged in favor of the skeptic. They are far more concerned with the manner in which a particular belief is formed – with the reliability of the mechanisms that produce it, for example, or with the law-like connections (if any) between the formation of one's beliefs and their truth.

In a thumbnail history of analytic epistemology the last three or four decades might fairly be described as the progressive triumph of externalism. Quine's naturalized epistemology provided a framework for a new conception of the relation of philosophy to empirical science, and within that framework various forms of externalism have flourished. At the dawn of the twenty-first century the dominant view seems to be that traditional

epistemology, with its insistence on *a priori* canons of rationality and its concern with the traditional skeptical problems raised by Descartes and Hume, is a hopeless enterprise and should be replaced by some form of naturalism.

This book is a sustained counterattack on behalf of traditional internalist epistemology. Our aim is threefold: to address some key criticisms of internalism and show that they do not hit their mark, to articulate a detailed version of a central objection to externalism, and to illustrate how a consistent internalism can meet the charge that it fares no better in the face of this objection than does externalism itself.

The first chapter addresses perhaps the most famous attempted counterexample in the history of philosophy. In a brief but profoundly influential paper published in 1963, Edmund Gettier argued that one can have a true, justified belief that *p* without having knowledge that *p*. Externalists have seized upon Gettier scenarios to argue that internalist notions like justification fail to provide us with the resources for a satisfying definition of knowledge. If they are right, internalists are trying to do epistemology with the wrong set of tools.

In response, we formulate and defend a purely internalist approach to the Gettier problem, arguing that an internalist foundationalist can give a convincing account not only of Gettier's original examples but of a wide range of more subtle and sophisticated scenarios. The heart of this account is a disambiguation of the notion of justification and a consequent analysis of the concept of knowledge that is particularly attractive since it does not require any concepts other than the time-honored notions of truth, internal justification, and belief. The internalist's tool kit, so to speak, is perfectly adequate to cope with Gettier-style scenarios.

The second chapter tackles what may be the primary objection leveled against epistemic internalists: that their account of knowledge, with its dependence on a purely internal conception of justification, is incapable of giving our beliefs the right sort of connection to truth. This is *prima facie* a serious charge. One of the aims of belief, perhaps the primary aim, is to believe truly – to get it *right*. If an epistemological position does not ensure that our beliefs at their epistemic best have any non-trivial connection to truth, then in what sense can it explicate any interesting form of justification or knowledge?

Whether internalism can answer this challenge depends on the sort of connection to truth one seeks. A major theme of this book is that there are two ways to conceive of such a connection: in terms of rationality, and in terms of success. Externalists focus on the latter sort of connection, by requiring (for example) that the belief-forming mechanisms of the knower be *reliable* in the sense that they produce, at least under normal conditions, true beliefs much more frequently than false beliefs. Call this an *extrinsic* connection to truth. According to externalists, what separates mere true belief from knowledge must be some type of extrinsic connection. And no

amount of introspection, no set of internal factors, will guarantee that our beliefs are connected to the truth in that way. If one sets up the epistemic challenge as a demand for an extrinsic connection to truth, then internalists have nothing interesting to offer.

But as we argue in Chapter 2, this mode of analysis is misconceived on multiple levels. The connection to truth can be seen in a clearer light when we turn our attention from belief-forming mechanisms to inference forms – the topic-neutral structures, whether deductive, inductive, probabilistic or explanatory, that give shape to our reasoning. If it can be shown that following such inference forms is *rational*, then beliefs arrived at by reasoning in accordance with them from unproblematic premises are genuinely probable or credible. It is this *intrinsic* connection to truth, according to internalists, that is at stake in classical skeptical arguments. The issue of the connection to truth, in the final analysis, is an insoluble problem only for those who have adopted the wrong view of what that connection should be – that is to say, for externalists.

Having disarmed these two threats to internalism, we turn in Chapters 3 and 4 to the examination of externalism itself. Here the discussion becomes rather complicated, but the guiding thread is that, according to externalism, one *cannot tell* whether one's beliefs are well-aimed at the truth. Here the threat of a profoundly troubling form of skepticism looms for those who adopt the externalist stance.

No sooner does an internalist raise this objection than an externalist will counter that the internalist is importing a question-begging notion of "telling" into the argument. If what is required for knowledge is (say) reliable belief production, then according to the externalist all that one needs to know that one's belief that p is reliably produced – to know that Rp – is to have one's true belief that Rp be *itself* reliably produced. Other externalists, such as Plantingians who use the concept of proper function, can make a similar move.

But from an internalist perspective this pushes the question further away: is that belief – that *Rp is reliably produced* – itself reliably produced? An infinite regress looms in which every challenge to a claim of reliable belief formation is answered in a way that produces another claim just as problematic. This is not the classical regress that shapes the controversy between foundationalists and their critics, however; it is a *metaregress*, a regress in which each question arises at a higher epistemic level where some putatively epistemic term has been iterated.

In the third chapter we explain this metaregress in more detail and explain how it illuminates a crucial difference between internalists and externalists. On an externalist account, it transpires that the metaregress can never be stopped because for any given statement, the epistemic defense of that statement at the metalevel will itself contain an empirical claim (about, say, the reliability with which the belief that Rp was produced) that is itself as much in need of defense as the original claim. But a careful internalist –

what we call an *armchair* internalist – does not fall prey to the same difficulty. For when internalism is properly understood, so we argue, the defense of an ascription of epistemic status *never* introduces new empirical information at the metalevel. This constraint has the consequence that epistemic principles are all necessary and *a priori*. Thus, although externalists are committed in principle to a position that engenders a very problematic regress, internalists are not similarly committed.

But is the metaregress really so problematic? In the fourth chapter we address the externalist attempt to bite the bullet and accept, in some cases revel in, the epistemic circularity endemic to externalism. Our suggestion, which may sound startling at first, is that the acceptance of epistemic circularity commits the externalist to the position that there is no such thing as knowledge. The argument hinges on what we call the modal principle, a preliminary version of which runs thus:

> MP: If it is in principle impossible to show decisively that S's belief that *p* is justified, then S is not justified in believing that *p*.

Through a series of approximations and refinements of MP we argue that externalists must accept a refined version of this principle on pain of arbitrariness, since without something of this sort they are not in a position to rule out evidently farcical epistemic principles. But if they are committed to even a modified and in some respects weaker version of MP, then their externalism tells against them: epistemic circularity will rule out the very possibility of knowledge. Externalists are not just skeptics, then, but super-skeptics – epistemic nihilists.

It becomes, at this point, an urgent question whether internalists have the resources to escape from the regress that catches the externalist. In view of the way that the metaregress argument is set up, this hinges on the question of whether the internalist's epistemic principles satisfy the modal principle; and this in turn depends on the nature of *a priori* knowledge. In the fifth chapter we stake out a modest claim, arguing merely that some *a priori* knowledge is analytic and that the epistemic principles needed by the internalist are of this sort. More daringly, we argue that a refined and expanded version of Locke's doctrine of intuition is correct: there are occasions on which we are infallible in our beliefs, and some of our beliefs about the interconnections of our concepts are cases of that sort. The combination of an analytic *a priori* account of epistemic principles and an intuitional account of the foundations of analytic *a priori* knowledge suffices to secure the internalist against the metaregress.

Externalists will surely retort that the security is purchased with *Scheingeld*. It is a philosophical commonplace that any position of the Lockean sort is open to two devastating objections: that it is inconsistent with human fallibility, and that we could never know for sure whether we were really "grasping" the relations of our concepts or were merely seeming to grasp

them. The former problem seems to rule out the sort of certainty promised by the internalist; the latter prevents him from taking any philosophical satisfaction from such certainty even if it exists.

Both objections, we argue, are mistaken. One may hold without inconsistency both that humans are frequently prone to fail and that there are some sorts of knowledge that are infallible. From the fact that we sometimes fail we may justly infer that we are sometimes not infallible, but it does not follow that we are not sometimes infallible. The latter objection involves a similar illicit inference. Certainly we sometimes have doubts about necessary truths – who can deny this? But it does not follow that we can never have a grasp of any necessary truths that leaves no room for doubt.

An account of *a priori* knowledge may be defensible in the abstract and yet fail to do the work one hoped for. In the final two chapters we come back to the internalist's focus on inference forms and sketch an account of deductive and inductive inference that is not vulnerable to the charge of epistemic circularity.

For deductive logic, the threat of epistemic circularity arises very naturally when we use deductive reasoning to demonstrate that our logical system has certain desirable properties such as soundness or completeness. Some externalists have argued that such proofs are epistemically circular but that this does not much matter. But this is demonstrably false. The sort of "justification" afforded by epistemically circular arguments provides us with no effective means of distinguishing valid from invalid inferences.

Logical intuition is necessary to ground our metatheoretic demonstrations. And indeed, without logical intuition there is no stopping the metaregress. But how can it be upheld in face of our fallibility in logical matters and the existence of rivals to classical bivalent logic? In part this challenge reduces to one we have already seen in Chapter 5, but the issue of deviant logics deserves consideration in its own right. We argue in Chapter 6 that the putative challenge of deviant logics is a semantic rather than an epistemic issue, that the only deviant logics worth taking seriously – and we do not deny that some of these systems may have intrinsic interest and useful applications – involve shifts in the meanings of certain concepts or the domains over which the deviant logics range. Seen from this point of view, however, "deviant" logics are innocuous; and in particular, their existence poses no challenge to the intuitability of elementary theorems of classical bivalent logic.

In the final chapter we apply some of the key distinctions developed in earlier chapters to the problem of induction. After reconstructing Hume's famous dilemma, we give an *a priori* defense of the *structure* of inductive arguments. By focusing on the inference form we demonstrate that it is possible, Hume notwithstanding, to give a cogent and non-circular vindication of the claim that under appropriate circumstances inductive arguments render their conclusions genuinely probable. The sense of probability employed is one that has to do with rationality rather than success; it offers

an intrinsic rather than an extrinsic connection to truth. But that is exactly what is required for a solution to the problem of induction.

Every philosophical work that engages with the current literature is structured partly by its choice of foils. In our book, that role is played primarily by three justly prominent contemporary epistemologists: William Alston, Alvin Plantinga, and Laurence BonJour. But BonJour's inclusion in this group is clearly anomalous. As a strong foundationalist and a staunch critic of naturalized epistemology and all forms of externalism, he is far closer to our position on nearly every issue than the others. If in this work we seem to be focusing almost exclusively on the residual areas of disagreement, it is largely because BonJour's neo-rationalism provides a more respectable foil for our theory of *a priori* knowledge than currently popular forms of neo-behaviorism – which, like him, we deplore.

Portions of this book were originally published elsewhere, though in each case such material has undergone revision (sometimes very extensive revision) during its incorporation. Older versions of Chapters 1 and 7 first appeared in *Journal of Philosophical Research* and *The Monist*, respectively. Chapter 4 contains material that first appeared in *American Philosophical Quarterly* and *Dialogue*. We are grateful to these journals for permission to use this material, and we give details of the original publications at the end of each of these chapters.

We have been fortunate in receiving useful critical feedback on various parts of this project. Evan Fales provided criticisms of our original Gettier paper that have influenced our revision for Chapter 1. We especially want to thank Richard Fumerton for comments on a draft of Chapter 2 and for extensive, helpful discussion of epistemic probability, the connection to truth, and other issues. His influence is apparent throughout Chapters 2 through 4, where our position and his are similar, though not identical. Tom Vinci and Steve Maitzen offered thoughtful criticisms of an earlier paper that helped us to improve our discussion of epistemic circularity. Aaron Cobb's work on internalism prompted us to clarify our position *vis à vis* higher-level requirements in Chapter 4. John Shoemaker raised the question treated in Chapter 5 about the relation between a correspondence theory of truth and a conceptual theory of analyticity. Henry Kyburg's influence on Chapter 7 will be obvious to anyone who knows his work. We are greatly indebted to him not only for his enormous contributions to the study of induction and probability but also for the solicitude with which he encouraged the writing of the original *Monist* article. Our sincere thanks to John Shoemaker for his initial work on the bibliography and to Jonah Schupbach who provided invaluable help completing the bibliography and creating the index.

1 Internalism and the Collapse of the Gettier Problem

Among challenges to internalism, the Gettier problem has pride of place. Four decades of discussion and analysis have produced no consensus on the problem itself, but the externalist response has been (not surprisingly) to tout it as evidence that internalism is seriously flawed. Alvin Plantinga, in particular, has argued that the Gettier problem is not just a minor conundrum requiring us to "tweak" the traditional TJB definition of knowledge but rather a sign that external factors regarding proper function are constitutive of the deepest and most important epistemic concepts.

For their own part, most internalists are understandably unwilling to define knowledge in general as requiring infallibility; to do so would preserve internalism at too high a price. But the introduction of a fourth condition involving causal grounding or reliable production mechanisms serves to weaken the impression that possessing beliefs and relying on them as reasons is of more than "folk psychological" importance.

In the rush to externalism, an alternative possibility for internalists has been largely ignored and, when discussed, has almost invariably been given short shrift. The heart of this solution is a clarifying analysis of what Gettier (and many others) took to be the J condition in the traditional analysis of knowledge. We will argue that the term 'justification' must be disambiguated. In one sense, it is equivalent to rationality; in another – the sense in which it is required for inferential knowledge – it involves an extension of the T condition.

We will argue that, while externalists have used Gettier as a motivation for the addition of various external causal requirements for knowledge, the externalist use of the Gettier problem has involved an ambiguity regarding what counts as "accidental knowledge." We will then consider a variety of attempted counterexamples to the solution we advocate, demonstrating that in each case the example fails and our analysis enables us to pinpoint the reason for its failure. Finally, we will discuss the effects of adding a T condition to crucial premises for the second sense of 'justification' and demonstrate that, far from pushing us to an externalist concept of knowledge, questions about the T component of the J condition actually collapse into the traditional epistemic regress argument.

Two senses of 'justification' and closure

In his seminal article, Gettier sets out to counterexample the traditional definition of knowledge as true, justified belief by exhibiting two cases in which a putative knower has true, justified belief but fails to have knowledge.[1] Before laying out the examples, Gettier notes two assumptions that are necessary to make the examples work. First, in the pertinent sense, S may be justified in believing that q without its being the case that q: justification does not entail truth. Second, if S is justified in believing that q, and q entails p, and S deduces p from q and accepts p as a result of this deduction, then S is justified in believing that p: justification is closed under known entailment.

Gettier goes on to set out two attempted counterexamples to the TJB analysis of knowledge, each of which has the following characteristics:

1 An individual S has evidence e which justifies his belief that q
2 S knows that q entails p
3 S forms the belief that p on the basis of his knowledge of the entailment relation from q and for no other reason
4 As it turns out, q is false
5 As it turns out, p is true.

Gettier maintains that p is a justified belief because of 1–3 and the principle that justification is closed under known entailment. But by 5, p is true. Hence p is a true, justified belief for S. Yet it seems intuitively obvious that it cannot count as knowledge; hence, true, justified belief is insufficient for knowledge.

Ironically, both Gettier's puzzle and the central insight required to resolve the difficulty were indicated by Bertrand Russell half a century prior to the publication of Gettier's paper.

> To take a very trivial instance: If a man believes that the late Prime Minister's last name began with a B, he believes what is true, since the late Prime Minister was Sir Henry Campbell Bannerman. But if he believes that Mr. Balfour was the late Prime Minister, he will still believe that the late Prime Minister's last name began with a B. . . . [He] may proceed to draw valid deductions from the true premiss that the late Prime Minister's name began with a B, but he cannot be said to *know* the conclusions reached by these deductions.[2]

As Russell points out, the obvious way to rule these cases out is to amend our definition of knowledge, and as a first step he suggests the following: "Knowledge is what is validly deduced from known premises."[3] Russell goes on to modify this definition in various ways; what he does not change, and sees no reason to change, is the emphasis on the importance of true premises in inferential knowledge.

The Russellian solution to the Gettier problem seems to face a difficulty right at the start. For clearly there is *something* epistemically significant about reasoning from beliefs for which one has reasons, even if it turns out that those beliefs are false. Indeed, to abandon that intuition would be to abandon internalism altogether. Russell's solution, with its stress on true beliefs, might appear to downplay the importance of rationality, and in that case it would be a non-starter.

But the objection rests on the unspoken premise that supporting beliefs play an important role only insofar as they produce knowledge. To undermine this assumption, we need to mark a distinction between two senses of 'justification.' Premises that provide justification$_1$ to *p* are (roughly) propositions on the basis of which S believes that *p*, which provide adequate grounds for rationally believing that *p*, and which are themselves held with a rational degree of credibility. The crucial premises of an argument that provides justification$_2$ are premises in this sense, but they must also be true. (We shall discuss the concept of a crucial premise in detail later in the chapter.) Transferring this analysis to Gettier's cases, S possesses justification$_1$ but not justification$_2$ for believing that *p*, since he grounds it crucially in *q*, which is false. And therefore, *contra* Gettier, S's belief that *p* does not satisfy the definition of knowledge.

Rather surprisingly, this solution has not gained wide acceptance, despite the fact that, in putting it forward, Russell solved the Gettier problem before Gettier. Besides Russell's anticipatory treatment, two articles in the mid-1970s advanced the core of this solution after Gettier. Robert Meyers and Kenneth Stern put forward a version of this solution (with explicit acknowledgment given to Russell) in which they mark the distinction by introducing "well-takenness" to indicate what we are calling "justification$_1$."[4] John Dreher, in a brief piece, distinguishes being "justified in believing" from "believing on good evidence" – where his being "justified in believing" refers to justification$_1$ and "believing on good evidence" indicates justification$_2$.[5] But in the face of a flurry of putative counterexamples the Russellian position has received no further public support, and Robert Shope's treatise on the Gettier problem gives it a fairly cursory dismissal.[6]

One consequence of the Russellian position that does not change despite varying terminology among its proponents is that the justification required for inferential knowledge (justification$_2$) is not closed under known entailment. Gettier explicitly invokes the premise that justification is so closed,[7] and it may be that the apparent counterintuitiveness of denying closure has prevented acknowledgment of the force of the solution. But as Meyers and Stern note, Gettier's article does not distinguish between the two senses of justification. When these two concepts are disentangled, we can argue that justification$_1$ is hereditary, i.e. closed under entailment known with certainty, although justification$_2$ is not.[8] When we realize that justification$_1$ isolates the distinctively epistemic nature of justification – i.e. rationality – it is clear that it *should* be closed under certainly known entailment, since

entailment preserves rational confidence. (See Chapter 2.) But since both senses of justification leave open the possibility that a proposition could be justified but false, it is obvious that justification$_2$ is not closed. One may rationally believe a proposition on the basis of a falsehood, if the falsehood itself is rationally believed, but the relation involved in justification$_2$ requires that the crucial premises of one's argument be truths as well.

We should note briefly here that, for the strong (incorrigibilist) foundationalist, it is not possible for a *foundational* belief to be rationally held but false, and (of course) foundational beliefs are exactly those that do not require *any* premises. Hence, a strong foundationalist will take the distinction between these two types of justification to be relevant only to inferred beliefs. Foundations, on this theory, have all the justification required for knowledge in themselves and are not subject to the possibility of Gettier undermining. We can think of this in various ways, and as long as we avoid confusion it does not much matter which we choose. We can say that justification$_2$ is required only for inferential knowledge, or we can say that justification$_2$ requires only that the crucial premises for a belief, *if there are any such*, must themselves be known. If one expresses the requirement in the latter terms the strong foundationalist would say that the foundations do satisfy the justification$_2$ condition because there are no rationally required premises for them. It is in part to avoid foreclosing the possibility of strong foundationalism that we have explicated the two senses of justification above in terms of *what sort of justification premises can provide* to a belief inferred from them, not in terms that might seem to apply to all beliefs a subject could hold. And as we shall see in the end, the Russellian solution to the Gettier problem does point to a need for foundations.

The disparity between the rationality of the subject in believing and the falsity of at least one important premise gives Gettier cases their distinctive characteristics: the subject's belief is rational while the conclusion is "true for the wrong reason" and therefore "adventitiously true." The requirement of true crucial premises for inferential knowledge together with the conferral of justification$_1$ upon the subject's belief addresses both aspects of Gettier scenarios.

The externalist use of the Gettier problem

Before responding to attempted counterexamples, we need to understand how and why the Gettier problem has been taken to lead to externalism. That it has had the historical effect of motivating a move to various naturalist and externalist epistemologies is beyond doubt. Once the TJB definition of knowledge was called into question, the idea arose that the much-sought "fourth condition" that must be added had something to do with the circumstances in the real world under which the subject formed his belief, perhaps even circumstances about which the subject had no beliefs or evidence at all. While it is possible to add a requirement that these real-world

circumstances be "right" on top of an ordinary internalist requirement for a good argument (or for internally accessible acquaintance with foundations), it was disturbing to traditional philosophers to think that some set of causes or circumstances – especially physical circumstances quite separate from the evidence available to the subject – might be indispensable to the subject's epistemic state. Traditional internalism involves a (sometimes unexpressed) requirement that at least any *empirical* truth that must obtain for the subject to have knowledge must be a truth the subject has some reason to believe.

William Alston is emphatic about the impact of Gettier upon the internalist-externalist debate. He implies that the Gettier problem made it very difficult to define "access internalism" because, he says, even self-styled internalists have now acknowledged that there is something other than truth that is necessary for knowledge to which we do not have even "relatively direct" access.

> 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.[9]

Philip Kitcher describes Gettier's argument as, historically, a quiet but important harbinger of the rise of naturalism that was to follow, important in part because it "called into question" the traditional enterprise of explicating justification.

> Psychology reentered epistemology quietly. A central problem in the analysis of knowledge takes for granted a conception of knowledge as justified true belief and seeks to provide an account of justification (foundationalist and coherence theories of justification being the main rivals). In 1963, a short article by Edmund Gettier called this enterprise into question by describing instances in which people have justified true belief but do not seem to have knowledge. Initial responses to Gettier's problem usually followed the apsychologistic orthodoxy, attempting to impose logical conditions on the subject's belief that would rule out the problematic examples as cases of knowledge. In the late 1960's, however, a number of authors proposed that a solution to Gettier's puzzling cases must lie in differentiating the causal processes

> that generate and sustain belief on those occasions where the subject knows. These generic approaches were articulated with the same kind of attention to detail that distinguished apsychologistic attacks on the Gettier problem. Yet, from a naturalistic perspective, their primary significance was their break with the apsychologistic tradition. Analyses of the concept of knowledge (and, later, of justification) . . . could take into account the processes . . . that causally generate states of belief.[10]

One of these early responses to Gettier was by Alvin Goldman, who in his seminal article, "A Causal Theory of Knowing," proposes that Gettier problems can be solved satisfactorily only by requiring an appropriate causal connection between the subject's belief that *p* and the fact that *p*. Goldman does argue in this article that it is not necessary for the subject to have any knowledge of this causal connection in the case of non-inferential beliefs (in which category he includes both perceptual and memory beliefs) and that the explication of the relevant causes is a job for the sciences. However, he retains a more traditional approach to inferential knowledge, envisaging in those cases a subject who believes that a particular causal process has taken place, which belief would be false if the situation were Gettierized. Indeed, he argues that the subject must "correctly reconstruct" at least the "important links" in the relevant causal chain to have knowledge based on inference.[11]

In a more radical move, Peter Unger advanced a theory on which justification is not even necessary for knowledge. Unger's substitute for justification, intended to cover Gettier cases as well as other putative counterexamples to the TJB analysis, is that it be "not at all accidental" that the subject's belief is true. Unger applies this notion of accident directly to questions of statistical reasoning and imposes a rather stringent notion of what it means for a belief's truth to be "not at all accidental." He even insists that when the odds in a physical set-up really do favor the truth of a subject's belief, the subject knows this to be the case, and the subject's belief turns out to be true, there may still be enough of a chance that the subject will be wrong that the belief does not qualify as knowledge. So, for example, if there are 15 black balls and 85 white balls in a bag of a hundred, if the subject knows this and believes that he will get a white ball, and if he does get a white ball, Unger still thinks of this as objectionably "accidental." But he holds that the better the physical odds get, the more legitimate it is to call the subject's belief a case of knowledge.[12]

The analysis of Gettier cases as "accidentally true belief" and the application of this analysis to an attack on internalism are best represented in more recent years by Alvin Plantinga. Plantinga is explicit about the importance he sees in the Gettier problem for the internalism/externalism debate:

> Gettier problems afflict internalist epistemologies, and they do so essentially. The essence of the Gettier problem is that it shows

> internalist theories of warrant to be wanting. What Gettier problems show, stated crudely . . . , is that even if everything is going as it ought to be with respect to what is internal (in the internalist sense), warrant may still be absent. The real significance of Gettier problems . . . is not that they are relatively minor technical annoyances that prevent us from getting a counterexample-proof analysis of knowledge; their real significance is that they show justification, conceived internalistically, to be insufficient for warrant. We should therefore expect that an externalist account such as the present account will enjoy a certain immunity to Gettier problems[.][13]

Plantinga rapidly dismisses the Russellian solution to Gettier as a mere "attempted repair" to the TJB definition. He argues that the "essence" of the Gettier problem lies not in the fact that the subject's belief is supported by a falsehood, but rather in two other factors. First, the subject's belief in these cases is merely "true by accident," and second, what is wrong in these cases concerns some "feature of the cognitive situation" that is external to the subject – for example, something in the "cognitive environment" that is not going right or some part of the subject's "cognitive equipment" that is malfunctioning.

As regards the second of these, Plantinga assumes without conspicuous argument that the subject does not need, rationally, to use *beliefs* about these aspects of the "cognitive situation" and whether they are "going right" as part of his support for accepting his conclusion. If such beliefs are necessary in Gettier cases, then it is arguable that it is the falsehood of the premise that is essential, not the failure to satisfy Plantingian requirements such as "proper function," "operation in the intended environment," and so forth. The characterization of what is wrong in Gettier cases as involving essentially things that are not "internal in the internalist sense" is therefore tendentious in that it draws attention away from the substantial extensional overlap there will be between a false premise analysis and Plantinga's own analysis. Both Plantinga's requirement that nothing be "going wrong" in the cognitive situation and his concept of what is external are so broad that one could subsume under them the falsehood of any proposition to which the subject does not have infallible access at time t. For example, suppose that the subject thinks, at t, that he had a visual experience as of an apple on the kitchen table at t-1. On the basis of this belief he concludes at t, without checking again, that there is an apple on the kitchen table. In fact, he had no such experience, nor does he have any other evidence that there was an apple on the kitchen table, but someone has in the meanwhile set an apple on the table. This is, of course, a classic Gettier case, and in it what is "wrong" is that the subject is mistaken in his memory belief regarding his private visual experiences. His belief is true for the wrong reason; it happens to be true because of a state of affairs about which he has no knowledge, while the premise he is actually using is false. One *can* call the falsehood of

his memory belief "external" just in the sense that at t the subject does not have indefectible access to it, and, the category of proper function being very broad, one can include a memory error about private visual experience under the heading of a "failure of proper function." But this problem in the cognitive situation is hardly "external to the subject" in the same sense as is, say, deception.

Plantinga does use an example about believing that there is a sheep in a field to assert that one could "reason directly" to the Gettierized conclusion without passing through the relevant falsehood (in this case, the proposition "that is a sheep"); we shall examine below several examples in which a subject is supposedly "reasoning directly" and not using a false premise. We will argue that such cases are really enthymemes if the subject is being rational at all. For the moment, however, it is enough to note that Plantinga's rhetoric on the "external" nature of Gettier problems is potentially misleading. Neither a list of classic Gettier cases nor Plantinga's description of what is wrong as "essentially external" will serve to show that his externalist analysis, as opposed to a false premise analysis, is correct.

For this very reason it is particularly helpful to examine in detail Plantinga's other point – that in every Gettier case the subject's belief is "true by accident" – and his use, in this context, of a scenario regarding barn facades. For here we come to an example that will not be analyzed as a Gettier case both by externalists and by advocates of the Russellian solution. This failure of extensional overlap between the two analyses helps, in turn, to pinpoint a confusion that besets externalist uses of Gettier problems.

Plantinga illustrates the notion of a belief's being "true by accident" using several well-known classic Gettier scenarios. For example, he considers Gettier's original scenario in which Smith brags and gives all sorts of evidence that he owns a Ford. S, who has no evidence at all about the whereabouts of Brown, believes on the basis of Smith's behavior not only "Smith owns a Ford" but also the entailed disjunction "Smith owns a Ford or Brown is in Barcelona." As it turns out, Smith is lying, but Brown really *is* in Barcelona. Similarly, S sees what appears to be a sheep in a field and concludes, "There's a sheep in that field." In fact, what he sees is not a sheep but a disguised wolf. His belief is true because there is a sheep in a different part of the same field which, however, is not visually accessible to him. It is fairly easy to see why one would say that S's belief in these cases is "true by accident."

But Plantinga also treats as "accidental" knowledge a scenario in which S is traveling in the country. Contemplating something that looks just like a barn, S believes, "That's a barn." In point of fact, it *is* a barn. However, unbeknownst to S, the inhabitants of the region have constructed highly realistic barn facades in a three to one ratio to the real barns, so many of the other things in the vicinity that look just like the thing he is observing at the moment are *not* real barns. Plantinga considers that all of these are Gettier cases and exhibit the same feature – that it "just happens" that S is

right: "In each of these cases it is merely *by accident* that the justified true belief in question is true."[14]

Undeniably, *something* about Gettier cases moves us to say that the conclusion is true "adventitiously" or "by accident." Plantinga, significantly, takes this concept of being right by accident in a broadly statistical sense:

> It *just happens* that Brown is in Barcelona, that there is a sheep in another part of the field, that what you are looking at is a barn rather than a barn facade. ... In each of these cases, the belief in question could just as well have been false. (As a matter of fact, that's not putting it strongly enough; these beliefs could *much better* have been false. There are so many other places Brown could have been; there are many more barn facades than barns there in southern Wisconsin; ... and so on.)[15]

On Plantinga's view, then, Gettier cases are bound together by the fact that someone frequently reasoning in that fashion under those circumstances would, at a minimum, be right no more often than he was wrong. Plantinga completes the move from Gettier to his own externalist epistemology by using Gettier cases to develop the notion of a properly functioning, truth-conducive module of a design plan operating in its intended environment. "[T]he first thing to see about Gettier situations is that the true beliefs in these situations are true by accident, not by virtue of the proper function of the faculties or belief-producing mechanisms involved."[16] Analyzing the Brown and the barn cases as failures of the "Principle of Credulity," a portion of the design plan that is supposed to operate in a particular type of environment, he argues, "[C]redulity is designed, we might say, to work in a particular kind of situation ... But when our fellows tell us what they think is false, then credulity fails to achieve the aimed at result."[17] All of this suggests that an "accident" in the sense relevant to Gettier cases stands in contrast to something which occurs by the (perhaps nomological) operation of a mechanism or system rendering the formation of true beliefs by that means in those sorts of circumstances usual or normal.

But a closer look reveals that Plantinga has conflated two different kinds of "accidents." Even if we consider only the three cases already described, there is something a bit odd about saying that the conclusion in the barn scenario is "true by accident" *in the same sense* that the conclusion is "true by accident" in the other two cases. (We will offer below a more direct argument that the barn scenario is different from classic Gettier scenarios.) Consider the wolf disguised as a sheep: the subject sees something that looks like a sheep and draws the conclusion, "There is a sheep in that field." But what he has seen isn't really a sheep. His belief, "There is a sheep in that field," just happens to be true or "is true by accident" in the sense that it is true for a reason of which he is entirely unaware – because of the presence of a sheep in some other part of the field – rather than being true for

the reason he is actually using – the presence of the animal he mistakenly identifies as a sheep. (The idea that the subject really does believe that what he sees is a sheep is nearly irresistible here, and we shall discuss below the problems attendant on insisting that the subject simply skips this belief in his reasoning.) It almost comes as a surprise, when hearing about the supposedly parallel barn case, to be told that the subject sees something that looks like a barn, believes that it *is* a barn, and, as it turns out, it is a barn. That's all there is to the scenario, except for another set of facts – which has no parallel in the sheep case – about the behavior of the inhabitants in the local region.

Similarly, in the case of Smith, appearances are deceiving. Smith is acting just as though he really owns a Ford, but he doesn't. The subject believes that what appears to be the case is the case – that Smith owns a Ford – and forms some other belief on that basis; it is the other belief ("Smith owns a Ford or Brown is in Barcelona") that one takes to have been Gettierized. That belief is "true by accident" again in the sense that it is true because of a fact about which the subject has no evidence – Brown's being in Barcelona – rather than because of the truth of a premise he is actually using – the proposition that Smith owns a Ford. But again, in the barn case, appearances are *not* deceiving: What looks like a barn is a barn, and yet we are supposed to believe that this is just like the other cases because of the falsehood of a proposition about the truthfulness of the inhabitants of the region (perhaps something like "the inhabitants of this region have not deliberately set out to deceive sightseers"), a *proposition that the subject is not contemplating or using in any way at t*. Does not the characterization of all of these as epistemically similar in that their conclusions are "true by accident" have about it the feel of sleight of hand? We should wonder from this comparison alone if something has gone wrong in Plantinga's analysis.

The fact that Plantinga's concept of "truth by accident" is faulty can be seen even more clearly if we modify the Smith case to make it actually parallel to the barn case. Suppose that S is a professor who has just taken up a post at a university in Smullyansville, where all of the native inhabitants, unbeknownst to S, are liars. The vast majority of the professor's students and colleagues are native Smullyanites. Smith, however, being from a different part of the country, does not have this overwhelming predisposition to lie; he is, in fact, as inclined to tell the truth as the majority of the other people S has known prior to taking up this position. Now, let us imagine that Smith, when he says that he owns a Ford, shows the papers of ownership, and so forth, is telling neither more nor less than the truth: he does indeed own the Ford. S believes him and infers, further, "Smith owns a Ford or Brown is in Barcelona." This belief itself is true, and it is true because of Smith's Ford ownership regardless of Brown's whereabouts.

This scenario is now very much like the barn scenario, but it is obviously quite different from the original case in which Smith himself is lying. Indeed, it is difficult to see why this should not count as a case of knowledge

on the part of S. Why should the mere fact that Smith is, so to speak, "surrounded by liars" make any difference to the question of whether it is possible to gain knowledge by listening to what Smith says? Smith does resemble the group of people from which S is presumably drawing his inductive inference regarding the general truthfulness of human beings, and Smith does own a Ford. Where has anything gone epistemically "wrong" in this entire situation?

One could say that S was in one sense "lucky" to be listening to Smith, since he could easily have been conversing with some other person nearby and hence believing falsehoods, but what difference does *this* sort of "luck" make to his having knowledge? It is quite different from the adventitiousness involved if his conclusion turns out to be true "for the wrong reason" – because of Brown's whereabouts, about which he had no evidence, rather than because of Smith's car ownership, which he was assuming in his reasoning.

Similarly, in the case of the barn, the real barn does in the relevant sense resemble the other barns from which the sightseer is presumably drawing his data. As in those cases, so here: a barn-like appearance to the observer is in fact caused by a real barn.[18] The observer's conclusion is not "adventitiously" correct in the way that it would be if, say, he were actually viewing a facade, concluded "There is a barn in that field," and (unbeknownst to him) there happened to be a real barn somewhere else in the field. That would be a true Gettier case; this is not.

Prima facie, the case in which Smith is lying should be treated differently from a case in which he is telling the truth. The barn scenario should therefore also be treated differently from classic Gettier cases. Even if Plantinga wants to use the barn scenario itself to argue for his proper function analysis of knowledge, it represents and promotes confusion to group it with Gettier scenarios and to claim that, as a group, they constitute evidence that internalism is incorrect because they all manifest the *same* epistemic problem.

If Gettier's original article had used "truthful Smith" and barn scenarios, it is doubtful that they would have had much influence on the subsequent development of epistemology. The intuition that S does not have knowledge in the original Smith case is simply much stronger than in the case where Smith is telling the truth but is surrounded by liars; in fact, we find nothing intuitive in the suggestion that the latter circumstance scotches S's knowledge. If nothing else, the resemblance of the barn case to that of truthful Smith surrounded by liars is reason to think that the subject in the barn case does have knowledge.

The claim that S lacks knowledge in these cases because he is "right by accident" is dreadfully vague, and attempts to flesh it out in terms of faculties and the like move into externalist territory where internalists – especially those who do not share the *initial* inclination to deny knowledge in barn scenarios – are not required to follow. Unger's attempt to specify a

statistical concept of what it means for the truth of one's belief to be "not at all accidental" suffers, as any similar attempt must, from the arbitrariness of the (on Unger's view very high) cut-off for the ratio of ways one could be right to ways one could be wrong. The arbitrariness should raise our suspicions that this is a false step. Why *should* we require for knowledge that the truth of one's belief not be "at all accidental"?

A clash of intuitions aside, a defense of the Russellian solution to the Gettier problem does not require us to argue that S in the barn case does have knowledge. We are strongly inclined to think that he does, because we do not see that any rationally essential premise of his argument is false and because we see no reason to accept a general "no accident" requirement on knowledge. We might answer the question, "Why do you think S in the barn case has knowledge?" by saying, "Because, under a very plausible set of epistemic conditions, S has true, justified belief, and the justification is of such a sort as to avoid Gettier counterexamples. What more do you want?" But even if others disagree about the barn scenario, what is more important for our purposes here is that the two Smith examples are very different, that the barn scenario is like the case of truthful Smith, and that the case of lying Smith is the real Gettier case.

By distinguishing the type of accident involved in classic Gettier cases – where the subject's conclusion is "true for the wrong reason" – from a *general* concept of an accident as something nomologically or statistically improbable, we remove an important link in the purported chain of reasoning from Gettier to externalism. It simply is not enough for externalists to say that their analysis of Gettier problems is better than an internalist analysis because they refuse to characterize as knowledge cases in which the subject's belief is "true by accident." We must decide what sort of accident is in question in a given example and, for different types of accidents, decide independently whether they preclude knowledge. Even a strong intuition that Gettier cases are not cases of knowledge need not move us to a general requirement that beliefs not be "true by accident." Any such broad fourth condition must be argued for independently of Gettier examples.

Intuitively satisfactory as the Russellian position is, its brief emergence as an answer to Gettier has been buried in alleged counterexamples that have never been systematically answered. These counterexamples range from cases that admit of easy explication to those that do not admit of one obvious analysis, and they can be separated into two categories. Although in Gettier's original examples S clearly believes the relevant falsehood,[19] one set of examples is intended to show that a Gettier problem *can* arise even when the subject is not relying on any false premise. These examples therefore imply that the Russellian solution is too lenient, that it would allow as "knowledge" beliefs that are in fact not known, and that because of a Gettier problem. A second (and much smaller) set of counterexamples purports to show that the Russellian solution is too restrictive, that it disqualifies some beliefs although the subject obviously does have knowledge.

Is the Russellian solution too permissive?

The counterexamples alleging a Gettier case where the Russellian solution does not detect one are numerous and, to some extent, miscellaneous. But we may divide them roughly into those that allege that the subject can reason in some especially "direct," careful, or clever way that avoids reliance on a false premise and those that allege that a Gettier case is present because of a causal problem either in the origin of the subject's belief or in the etiology of the event that makes true.

Many counterexamples of the first type rely on the assertion that S can reason *directly* to a conclusion. We shall argue that in such cases, either a) S is not being rational in believing the conclusion, and therefore there is no Gettier case (since S is not justified in any sense at all), or b) S is indeed relying on the false proposition, and appearances to the contrary arise only because it is so easy to imagine S as reasoning enthymematically, not making all of his premises explicit.

The first relatively simple case of this sort comes from Keith Lehrer:

> *Mr. Nogot (Non-Discursive Version)*: A pupil in *S*'s office, Mr. Nogot, has given *S* evidence *e* from which *S* directly infers *p*: 'Someone in the office owns a Ford.' But unsuspected by *S*, it is Havit [who is also in the office] who owns a Ford and Nogot does not.[20]

This claim that S can "directly infer" *p* and avoid a false premise is exactly like Plantinga's claim that one could "proceed directly" to the conclusion that there is a sheep in a field without believing "that is a sheep."[21] (In Plantinga's example, as we have already discussed, what S sees is not a sheep, but there is indeed a sheep elsewhere in the field.) But it is very difficult to grasp how those who put forward such examples intend S to be reasoning so that he is at least rational. If S is not rational, of course, there is no Gettier case, since reasonable belief is required to satisfy any version of the J condition and all Gettier cases purport to satisfy the TJB conditions.

It is possible that Plantinga is assuming (as, perhaps, is Lehrer) that all inference must be explicit inference. This assumption is indicated tellingly in Plantinga when he uses the word 'first': "Modify the sheep case so that you don't first form the belief that *that is a sheep* but proceed directly to the belief that there is a sheep in that field."[22] But no internalist is required to claim that inferential support relations occur in a temporal series. Indeed, evidential support ought to be construed synchronically. And in that case, a great many premises will be implicit. An internalist need not be committed to the claim that a subject must have all of the assumptions underlying his conclusions explicitly before his mind.[23] Furthermore, if there is no implicit premise to the effect that the creature in front of S is a sheep, it is hard to see how he can be reasonable at all in believing that there is a sheep in the field, assuming that he has no other evidence relevant to that proposition.

Similarly in the Nogot case, what exactly is S's argument? If we take evidence *e*, given by Mr. Nogot, to be some sort of behavior such as stating that he owns a Ford or showing ownership papers,[24] then the argument looks like this:

1 Mr. Nogot has behaved before me in such-and-such a way, indicating that he owns a Ford.
2 Mr. Nogot is in my office.

Therefore,

3 Someone in my office owns a Ford.[25]

It is the step from 1 and 2 to the conclusion, 3, that Shope (summarizing Lehrer) calls "inferring directly." But surely it should more accurately be called *either* "inferring enthememically" or "inferring irrationally." For as stated, the argument is insufficient to probabilify the conclusion. Either S is at least implicitly invoking the premise:

1′ Mr. Nogot owns a Ford,

or he simply does not have a good argument and hence does not possess justification in any sense.

Another example from Lehrer, the case of "the Clever Reasoner," can easily be analyzed in a similar fashion.

> *The Clever Reasoner*: A pupil in *S*'s class, Mr. Nogot, has given *S* evidence e' which is sufficient to justify *S*'s believing *q*: 'Mr. Nogot owns a Ferrari,' but *S* lacks evidence bearing on *r*: 'Mr. Havit owns a Ferrari.' The teacher is not interested in who the Ferrari owners in his class may be but only in whether it is true that *p*: 'Someone in the class owns a Ferrari.' The teacher reasons that although the evidence supports *q*, there is at least the possibility that someone else in the class owns one, and, hence, that it is safer to accept the more general statement, *p* than to accept *q*. Accordingly, *S* accepts *p*. Mr. Nogot again has been shamming, but *r* and *p* are true.[26]

Once again, laying S's argument out in standard form shows it to be invalid as described:

1 Mr. Nogot has behaved as if he were a Ferrari owner.
2 Mr. Nogot is a student in my class.
3 It is possible that some other student in my class owns a Ferrari.

Therefore,

4 Someone in my class owns a Ferrari.

Since the example makes it quite clear that S refuses to accept a premise such as 1′ above (in this case, "Mr. Nogot owns a Ferrari") and that he has no evidence at all bearing on the truth of *r*, this appears to be all there is to his argument. Premise 3 is no help at all since, if we are simply talking about *possibility*, it is equally *possible* that no other student in the class owns a Ferrari. S therefore is not justified in any sense, and the case is not a Gettier case.

The phrase 'there is at least the possibility' is in fact rather vague, and it could be interpreted to mean that S thinks it actually *plausible* or *probable* that some other student in his class owns such a car. But even then, S cannot get down to the specific level of believing the conclusion of the argument from purely epistemic considerations about plausibility; he must be willing to go farther and believe that some other student in the class *does* own a Ferrari. The trouble with this analysis is that we have been told that S lacks evidence bearing on Havit's car ownership. Yet if this is true, then S must in fact not have any evidence bearing on the car ownership of students other than Nogot (since such evidence would be automatically relevant to Havit's car ownership), and in that case S cannot be reasonable in accepting "Some other student in my class owns a Ferrari." So in that case as well, S is not rational.

A subtler attempt to argue that S can simply skip false premises comes from Richard Feldman.

> *Mr. Nogot (Feldman's Version)*: This case is similar to [the original Nogot case], except that *S* does not arrive at belief in *p* by relying on considerations about *q* ['Nogot owns a Ford'], but instead by relying on belief in a true existential generalization from his evidence, of the form: 'There is someone in *S*'s office who has given *S* evidence *e*.'[27]

This is a rather cryptic statement (and Feldman's original version is more explicit only in that it gives more details of evidence *e* and states that Nogot has been a trustworthy person in the past),[28] but the argument it indicates can be laid out in what appears to be a convincing form like this:

1 Mr. Nogot, who is in my office, has behaved as if he owned a Ford.

Therefore,

2 Someone in my office has behaved as if he owned a Ford.

Therefore,

3 Someone in my office owns a Ford.

Since Nogot is the only person who has "given S evidence *e*" (or "behaved as if he owned a Ford"), premise 1 is required to justify, the "true existential generalization," 2. Since the phrase 'someone in my office' occurs in both 2 and 3, it might appear that the move from 2 to 3 is an uncontroversial inference requiring at most an intermediate (and, we may grant for the sake of the argument, true) premise connecting behavior in general with Ford ownership in general – e.g. "Given what I know, most people who behave as if they own Fords do own Fords." But this appearance is misleading. In the conclusion, 3, the word 'someone' must indicate an entirely unspecified existential quantifier in order to make 3 a true statement just in case *anyone* in the office (Havit, for example) owns a Ford. (If 3 were not true under these circumstances, this would not be a Gettier case.) Therefore, premise 3 cannot mean "The person in my office who has behaved as if he owns a Ford owns a Ford" (in which case it would be false) but must mean simply "There exists some person who is in my office and who owns a Ford."

But therein lies the problem. The second existential generalization cannot reasonably be inferred from the first without an *intermediary* premise such as

2′ The person in my office who has behaved as if he owns a Ford does own a Ford.

The first existential generalization (*ex hypothesi*) is supported only by evidence that a specific person (in this case, Nogot) had behaved as if he owned a Ford. Therefore, S cannot reason to a second existential generalization unless he reasonably believes, at least implicitly, that this same individual owns a Ford.

The point is important and deserves elaboration. In a natural deduction system one needs to work from a specific statement, using constants, when inferring by EG an existential generalization from the internal content of an earlier line. The rule for existential generalization, introducing the existential quantifier, can be applied only to a statement of the same form which uses a constant.[29] The only way to fill out Feldman's enthymeme validly without including *explicitly* the claim that Nogot owns a Ford is to introduce probabilities into the argument without paying sufficient attention to the constraints on their use. It is this casual treatment of probabilities that disguises the actual structure of dependence in the argument.

In order to see this clearly, we need to introduce some symbolic tools for the advocate of Feldman's counterexample to use. Let B stand for the predicate " ... has behaved as if he owns a Ford" and let F stand for the predicate " ... owns a Ford." Let p be a sentence-forming operator on sentences, to be interpreted "It is (epistemically) probable that ... " To make Feldman's case as strong as possible, we will give him a simplistic rule of probabilistic detachment (PD): from $p\varphi$, infer that φ, taking as premises all of the premises of the former line. We can now reconstruct his proof thus:

{1}	1	Bn	[Premise]
{1}	2	(∃x)Bx	[EG from 1]
{3}	3	(x)(Bx ⊃ *p*Fx)	[Premise]
{4}	4	Ba	[Premise for ES]
{3}	5	Ba ⊃ *p*Fa	[US from 3]
{3,4}	6	*p*Fa	[MP from 4,5]
{3,4}	7	Fa	[PD, 6]
{3,4}	8	(∃x)Fx	[EG from 7]
{1,3}	9	(∃x)Fx	[ES from 2–8]

The plausibility of Feldman's example rests on the claim that 'Fn' – the falsehood involved in the case – does not appear explicitly on any line of this proof. But is that the only relevant issue? Surely the oddest feature of the proof is the circuitous invocation of 'Ba' on line 4 when there is a perfectly good specification of the same form available already on line 1. It is obvious that existential specification is used only to avoid an open statement of 'Fn,' as there is a straightforward 6-step proof which does not use ES but does explicitly invoke 'Fn.'

To put a finger on the problem, we must return to premise 3 which claims that *anyone* who behaves in such-and-such a fashion probably owns a Ford. What could justify such a strong assertion? On reflection it is clear that this premise is not even derivable from the strong assumption

A: (x) P(Fx|Bx & K) > r,

where r is some high value in the interval [0, 1]; for probabilities (when they represent something of epistemic rather than of merely psychological significance) are always sensitive to background information, and while **A** makes reference to K, there is no background information expressed or presupposed in premise 3. In fact, the attempt to lay out an essentially non-deductive argument as a deduction with an uncertain (but probabilistically detachable) premise fails precisely because there is no way to represent the fact that K is all of the relevant background without stepping outside of deductive logic and casting the dependence as a conditional probability.[30]

If we shifted to the language of probability in order to take K into account, could the argument be reconstructed? The way that the scenario is set up we are invited to assume that for some tolerably high value of r, P(Fn|Bn & K) > r. Since 'a' and 'n' are both names, one might try to substitute one for the other and conclude that P(Fa|Ba & K) > r as well. But this is hopeless; for in the end the conclusion on line 9 rests not on Ba but on Bn, and *ex hypothesi* the subject has no evidence regarding the Ford ownership of anyone in the office besides Nogot. There is no way to make the argument work without routing it through a premise which attributes Ford ownership to an individual who has behaved as if he owned a Ford (as in 2′, above, and as in premise 7 of the argument given). But Nogot is the only *actual* person about whom S has any such evidence.

We can give this intuition a precise formulation by noting that, relative to the background information clearly presupposed in the problem, ±Fn *screens off* Bn with respect to (∃x)Fx:

$$P((\exists x)Fx|Fn \,\&\, Bn \,\&\, K) = P((\exists x)Fx|Fn \,\&\, K), \text{ and}$$

$$P((\exists x)Fx|{\sim}Fn \,\&\, Bn \,\&\, K) = P((\exists x)Fx|{\sim}Fn \,\&\, K).$$

This is a formal way of representing the commonsense fact that the relevance of Nogot's behavior to the question of whether someone in the office owns a Ford turns critically on the question of whether *Nogot himself* owns a Ford. Thus the probabilistic reconstruction of the Feldman case reveals something epistemically significant that is concealed in the deductive version of the proof: if we cannot run the *Modus Ponens* of lines 4–6 through Nogot, we cannot run it at all. It is in this sense that 'Fn' is a necessary implicit premise for the attempted justification of line 9, and its falsehood vitiates Feldman's attempted counterexample.

The attempted causal counterexamples to the Russellian analysis illustrate particularly well the one lesson that all epistemologists can learn from the Gettier problem: unargued assertions to the effect that S obviously does have knowledge or (more often) obviously does not have it are often highly controversial. We have already shown in our analysis of the barn scenario that internalists should be quite willing to insist that those who make up some scenarios are just wrong, that S does indeed have knowledge, and that the rather convoluted sets of circumstances that are supposed to vitiate his knowledge do nothing of the sort. The causal examples that follow show something similar: it is not obvious that "some causal problem" in an epistemic situation always undermines a belief's status as knowledge, and the more complex a scenario becomes, the more difficult it is to access any clear intuition regarding whether the subject has knowledge or not. Furthermore, scenarios involving causal mishaps are often, in virtue of their very complexity, quite different from Gettier's original, straightforward case and may not be Gettier cases at all. If they are not, then, as with the barn scenario, it is important that we classify them separately. Even if it still seems to many that there is no knowledge in these scenarios, they may not properly be regarded as part of a cumulative case against internalism arising from the Gettier problem.

Brian Skyrms sets out one of the most interesting causal examples used against a false premise analysis of the Gettier problem. Shope states that this example "does not involve either actual belief in false propositions or their employment in inference."[31]

> *The Pyromaniac*: Striking a match, S infers that it will light directly from S's knowledge that it is a dry match of a brand ("Sure-Fire" matches) that has often and always lit for S when dry and struck.

> However, unsuspected by S this one cannot be lit by friction because of impurities and is going to light only because of a burst of rare Q-radiation.[32]

Skyrms himself calls this a causal counterexample to the TJB analysis of knowing but says that it is *not* a Gettier case,[33] but Shope includes it as a Gettier case because of an observation by Marshall Swain that S is justified in believing various propositions such as *q*: "This match is like previously struck Sure-Fire matches in all respects relevant to ignition."[34]

The exact analysis one gives this example will depend upon what type of induction one permits in this case. There are two different possible analyses, but in neither of them is this the counterexample Shope believes it to be. On the one hand, one may believe that S in this situation can make a rational induction involving no implicit assumptions about underlying causal mechanisms. In this case, S's reasoning would look (roughly) like this:

1 m/n of the Sure-Fire matches I have struck have lit.

Therefore,

2 Approximately m/n Sure-Fire matches will light when struck.
3 The match I now hold is a Sure-Fire match.

Therefore (with confidence approximately m/n)

4 The match I now hold will light when I strike it.

Let us assume that S's evidence does not indicate that this match is any different from the other Sure-Fire matches he has encountered with respect to lighting when struck, or at least that he has no evidence that makes it *less* likely than those other matches to light when struck,[35] and that m/n is a high ratio. Suppose, then, that this is a good inductive argument in the context of S's total evidence. In that case, it seems quite irrelevant to point out that S is justified in believing *q*, a statement about the underlying structure of the match in his hand and the similarity of that structure to that of Sure-Fire matches he has struck in the past. Furthermore, if this argument makes it rational to believe its conclusion, we may well think that S also has knowledge, regardless of the impurities in this particular match or the existence of Q-radiation. Hence, in this analysis, there is no Gettier case.

Shope wants to say that *q* is relevant because, in the terms of his own attempted solution to the Gettier problem, *q* (like all relevant falsehoods in Gettier cases) is part of a "pseudo-justification-explaining chain" which would "explain" the justification of *p* for S if *q* were not false.[36] But if *q* is a part of the explanation for the (apparent) justification of *p*, then it must not

be possible for S in this situation to make a justificatory prediction without the use of *q*. If *q* is relevant, it must be relevant because it is required to justify the inductive conclusion, "The match I now hold will light when I strike it." In that case, S's argument looks roughly like this:

1 Every other time that I have struck a Sure-Fire match, it has lit.
2 The match I now hold is a Sure-Fire match.
3 The match I now hold is like previously struck Sure-Fire matches in all respects relevant to ignition.

Therefore,

4 The match I now hold will light when I strike it.

The false premise, of course, is 3.

Although an advocate of the Russell solution to the Gettier problem can present his interlocutor with this either/or (either a false premise is required or S has knowledge), the pyromaniac case is not trivial. It raises an extremely complex problem concerning what – if any – causal premises about past observations are required for a subject to make rational predictions. On the one hand, it seems incorrect to require for induction that we have beliefs – still less, that we have detailed beliefs – about the various causal factors underlying our past observations, in part because those causal factors may not form anything like a natural kind. The Sure-Fire match company may have changed its manufacturing process during the time S has been using the brand, so that the ignition even of the matches he has observed so far has been brought about by a somewhat varied match design. How much "like" other observed matches must the current match be in order to make *q* a true proposition? And how much *alike* must the observed matches have been in order to make it a meaningful proposition?

To take an even more difficult example, a man may see a dog in the yard of a particular house every day when he walks by at a certain time. But on one day, the causal history behind the dog's presence may include the fact that he was deliberately let out into the yard, while on another day he found an open window and went out, and on still another day he became separated from his master while on an outing and found his way back to the yard alone. In all of these cases, it is at least true that the dog lives at that house; but even this might change, if the dog were sold to a neighbor and returned to his old yard when no one was watching. It is almost impossible to find a single nexus of related causes underlying the dog's presence in the yard in all of these circumstances, and it is implausible to suggest that the man who sees him every day must believe something about such a causal nexus in order to reason inductively that he will see him the next day. The belief, "That dog has *something* to do with that yard" is so vague that it is difficult to believe it to be a rationally required part of S's argument.

But this argument is not entirely decisive regarding Skyrms's example. For many adults do know something about the process of match manufacture. They know, for example, that matches light because match heads are coated with substances that burst into flame by friction when the match is struck. To put the matter more pointedly, they know as an empirical matter that this is the *only* reason that matches light upon striking and that an uncoated match or one coated only with a non-flammable substance will not light when struck under any normal set of circumstances. There is some plausibility to the claim that, once one has this sort of causal knowledge, it is pretending ignorance to disregard that knowledge in setting one's future expectations. If this is correct, it seems at least true that S's degree of rational confidence in the proposition that the next match will light must result from routing his inference through a belief somewhat like q, such as

> q': This match has been coated with a sufficient quantity of an inflammable substance that friction against the side of a matchbox will cause it to light.

This proposition does not suffer from as much vagueness as the proposition about the dog, and if supported it would indeed help S to have a high degree of rational confidence in his expectation as to the match's lighting. In fact, he could use a smaller set of past observations of actual match-lighting while maintaining high confidence in his prediction if he learned some details of the process of match manufacture and made a supportive induction that those details would apply in the case at hand. An intuition (such as Shope's) that a premise like q or q' is in some way relevant to S's justification probably arises from the inclination to say that a well-informed adult must make that sort of causal assumption in order to have a rational expectation of the match's lighting. But if he must, then S as Skyrms portrays him is not justified in any sense at all, and again the case is neither a Gettier case nor a counterexample to the TJB analysis of knowledge.

Paradoxically, then, it may well be that the more one knows the more one needs to know in order to have further justified$_1$ (that is, rational) beliefs. A child, to whom all match-lighting is akin to magic, who has no theory whatsoever about the underlying causes of match-lighting, might require fewer justified causal premises in order to have a rational expectation that a new match will light. Assuming that induction is rationally legitimate at all (see Chapter 7), a person under those circumstances might well be able to make an inductive inference that the next match will light from a large sample of past lightings without supportive premises as to the constancy of particular causal factors in the next match. Indeed, for all he knows, perhaps all past matches have lit because of Q-radiation! And if S is in that evidential situation, he is justified in both senses and has knowledge.

The Skyrms example, then, raises questions that are not easily answered about what premises are required for rational predictive inference. But it

does not thereby undermine our solution to the Gettier problem. Indeed, the Russellian solution permits one to explain with precision the questions raised by the example.

J. Gregory Dees and John Hart have also attempted to counterexample the Russellian solution by alleging that we can conceive of a case in which S's conclusion is true, S is rational in believing that conclusion, his argument has no false premises, yet he fails to have knowledge because of a Gettier type of causal "accident." Dees and Hart's cases involve the claim that something has gone wrong in the causal antecedents of the subject's belief.

> Suppose that a student, Ms Right, reads in her philosophy text that *p*, "Spinoza was born in 1632." On the basis of this she infers that *h*, Spinoza was born in the seventeenth century. According to Meyers and Stern, Ms Right would know that *h*.
>
> But compare the similar case of Ms Print, who like Ms Right, reads in her text (a different text) the same information about Spinoza, and forms the same belief that *h*. Her case, however is not so simple. The author of Ms Print's text was mistaken about the date of Spinoza's birth, and sent his manuscript to the printer with the date as 1732. The typesetter who had been working overtime and was weary accidentally picked up a "6" rather than a "7," thus "correcting" the author's error. Consequently, it is purely accidental that Ms Print came to believe *h*, even though her evidence was true and acquired in an accredited way.[37]

The analysis of this example is not quite as obvious as Dees and Hart take it to be, for reasons broadly similar to those applying in the match-lighting case. It is clearly not necessary for a subject to have a detailed causal knowledge of the process of book production (and especially, the process of the production of *this* book) and of its reliability at every stage (from the trustworthiness of the author to the accuracy of each typesetter and copyist) in order to be rational in believing what he reads in books. And even the most committed externalist will not want to say that the slightest aberration from perfect accuracy in the process of writing and producing a book renders it impossible to obtain knowledge by reading the book. This raises at least some question as to when a belief about a statement in a book is "Gettierized." For example, if the author knew the date of Spinoza's birth and wrote it down correctly, if one copyist in the process copied it incorrectly as 1732 and a different copyist (also without knowing the actual date) miscopied that date and changed it back to 1632, would this Three Stooges scenario Gettierize Ms. Print's belief about Spinoza? The answer is, to put it mildly, unclear. There is something "accidental" about the correctness of the statement in the book, but the accident may simply be of the statistical sort discussed above – if this sort of thing happened often enough, beliefs

acquired from reading books would not often be true – rather than a genuine Gettier accident in which the belief is "true for the wrong reason." If this describes a genuine Gettier case, then the relevant proposition that turns out to be false would have to be something quite strong such as:

> q'' The sentence in my book deliberately and successfully reproduces a sentence expressing the intentions of a person knowledgeable about the date of Spinoza's birth.

The case Dees and Hart actually imagine is more obviously a Gettier case than the comedy of errors just constructed, since in their example the author himself is mistaken, and since (*pace* postmodernist theories) people reasonably believe what they read in books because they take the authors of books to know (more or less) what they are talking about. It is therefore plausible that, as in the pyromaniac case, some sort of belief about the causal history of the particular book forms a needed premise for Ms. Print's rational belief. For example, both Ms. Print and Ms. Right are, we may assume, justified$_1$ in believing

> q''' The sentence in my book expresses the intentions of a person knowledgeable about the date of Spinoza's birth.

Since it seems that both Ms. Print and Ms. Right need to be basing their belief about Spinoza at least implicitly on an assumption such as q''', Ms. Print has justification$_1$ but not knowledge, and Ms. Right (since we assume that all of her necessary premises are true) has knowledge.

From the examination of these attempted counterexamples, a consistent pattern emerges. In any case where the truth of S's conclusion is clearly adventitious in the classic Gettier sense of being "true for the wrong reason," the only question is how well the false premise has been hidden. While there are cases where it is not obvious *whether* there is a Gettier case at all, this is because it is not clear that a false proposition really is a necessary assumption for S to have a rational belief. Wherever a clear Gettier case exists, a sufficiently careful examination will uncover the lurking false premise. In many unclear cases it is even quite easy to identify a falsehood that *may* be relevant to S's justification, and the question then is only whether it really is epistemically relevant. We can therefore vindicate the Russellian analysis against the charge that it is too permissive.

Is the Russellian solution too restrictive?

Dees and Hart also criticize the Russellian solution for the opposite fault, arguing that it is too restrictive, that is, that it would require us to designate beliefs as not known when they clearly are known. This criticism requires careful analysis, and to give it the attention it deserves, we need to

make explicit an important concept that we have already mentioned and used implicitly: the concept of a "necessary" or "crucial" assumption. For purposes of analyzing Gettier cases, we may think of a crucial assumption q as a belief that is itself rational for S to hold on the basis of its own independent evidence and that is also the conclusion of a line of reasoning that is, in the epistemic context, necessary for the rationality of S's belief that p. So:

> q is a crucial premise of S's argument for p if a) S is justified$_1$ in believing that p, b) S is justified$_1$ in believing that q independent of p, c) S is basing his belief that p on q, and d) given S's other premises and their rational credibility levels, if S were not thus justified$_1$ in believing that q or were not basing his belief that p on q, S would not be justified$_1$ in believing that p.[38]

The concept of a crucial premise is especially useful in discussing cases where the Russellian solution is accused of being too restrictive. Dees and Hart first describe a case raised by Lehrer in which S bases his belief that p on the conjunction of q and some other proposition, either of which entails p. However, q is false, so the conjunction on which p is based is false.[39]

Meyers and Stern's response to this attempted counterexample is to say that S must realize that the other proposition is in itself a sufficiently good reason for believing that p, so that he would fall back upon the other proposition were he to discover that q is false.[40] Dees and Hart attempt to argue that this is too strong a requirement with a further example. They imagine that a research chemist, Wilma, has a false belief about justification that entails that her set of actual experiments is not large enough to justify her conclusion that two chemicals mixed in certain quantities yield a third chemical. Wilma believes that she needs five thousand repetitions of a given experiment to confirm a result but, unbeknownst to her, several of her attempted repetitions have been vitiated by a switch in chemicals. Wilma has as a premise that she has performed 5,000 actual repetitions of the experiment, but this premise is false. Dees and Hart build into this scenario the counterfactual claim that Wilma *would*, even if it were pointed out to her that she did not perform all 5,000 repetitions of the experiment, continue to believe her conclusion on the basis of the other experiments alone – in violation of her own epistemic commitment. Under these conditions, they maintain, Wilma clearly has knowledge. Since this result appears to contradict Meyers and Stern's principle regarding the subject's need to "know" that the remaining evidence is rationally sufficient, Dees and Hart conclude that the Russellian solution is too restrictive.[41]

The relationship between a subject's (false) beliefs about justification and his knowledge of the propositions to which those false epistemic beliefs apply is far more complicated than Dees and Hart seem to realize. Some philosophers maintain that a subject like Wilma does *not* have justification under the circumstances described.[42] But even if Wilma has knowledge,

Meyers and Stern need not have required true positive knowledge about justification on the part of the subject. It is fairly clear that they are trying to analyze the concept of basing a belief on some particular premise and to see how such basing affects the importance of the other premises one is using, but it is not at all obvious that basing requires positive knowledge about what is rationally required for justification.

We can therefore weaken Meyers and Stern's requirement while retaining its essential concern with basing: if the falsehood is not to Gettierize the conclusion, the falsehood must not be a crucial premise of the subject's argument. But the concept of a crucial premise involves no commitment on the question of whether the subject must have true positive beliefs about the nature of his own justification.[43]

Here the somewhat expanded Russellian solution respects the concern that we analyze the knowledge and rationality of *subjects* given the way that those subjects are inferring, not the purely abstract connections among propositions "floating around" in a subject's belief set but not used to support his conclusion. The notion that justification concerns only such abstract connections, aside from considerations of basing, is criticized by Hilary Kornblith as the "arguments on paper thesis,"[44] and we have argued elsewhere that internalism is not committed to this thesis.[45]

Dreher obliquely refers to the possibility that the falsehood is not a crucial premise and that its falsehood would in that case not vitiate knowledge by saying that an argument must have true premises (in order to provide knowledge) "unless there are extraneous false premises."[46] Similarly, in many statements of classic Gettier cases, it is explicitly stipulated that S has no other knowledge relevant to the conclusion, and we have mentioned the same requirement in discussing Plantinga's sheep case. The question, then, for attempted counterexamples in which S bases his belief on a conjunction is whether he bases it crucially on the conjunction itself or also on the conjuncts separately. In Lehrer's example, the subject infers his conclusion from a conjunction, either conjunct of which entails it. Here it is very plausible to imagine that a subject would be basing the conjunction itself on the separate conjuncts and on the evidence for them, and in that case it makes sense to say that the conclusion is based on the separate conjuncts and does not depend crucially on their conjunction. If in some odd context the subject is *not* basing his conclusion on the conjuncts but only on the conjunction, and if he does not have sufficient other evidence that does not suffer from Gettier problems, the falsehood of the conjunction does Gettierize his conclusion. But the rejection of the "arguments on paper thesis" indicates why that result is not implausible: justification is in part a matter of a subject's basings. It does not arise solely from abstract relations among propositions.

In the case of Wilma, the question is whether she bases her conclusion (that the two mixed substances yield the third) on the separable premises stating that each of the other experiments has been correctly performed.

Since Dees and Hart explicitly indicate that she *is* basing her conclusion on a sufficient number of premises for rationality, which premises are also true, the Russellian solution itself does not rule out her having knowledge. While other epistemologists may want to add further requirements – such as the requirement that the subject not have a belief that entails that his inference based on the other premises alone is or would be irrational, or that the subject must have a justified belief that his inference based on the other premises is or would be rational – these requirements all go beyond the core of the Russellian solution. Thus the second sort of counterexample also fails, and the Russellian solution proves resilient enough to meet the charge that it is too restrictive.

The collapse into the regress

We take it as axiomatic that on any reasonable account, truth is non-epistemic; our believing, wishing, hoping or fearing that *p* is not definitional of *p*'s being true. Such a condition seems the least that can be asked for any theory of objective truth. But when this stricture on truth is connected with the Russellian account of knowledge advocated above, an interesting worry arises. It seems that the set of conditions for knowledge, at least for inferential knowledge, now contains *two* non-epistemic components, the T condition on knowledge and the T requirement for crucial premises in the justification$_2$ condition. What wider implications does this fact have for the theory of knowledge?

Perhaps surprisingly, our argument indicates that there are few or none. Far from requiring a capitulation to externalism, the Gettier problem merely requires a disambiguating of the J condition. At that point, justification$_1$ emerges as by far the more interesting concept, since justification$_2$ is only justification$_1$ with the added requirement of truth for crucial premises (if there are any such). We can see this asymmetry between the two senses of justification in one of the most interesting unresolved questions raised by our solution to the Gettier problem – the role played by causal premises in non-deductive arguments, e.g. in the match-lighting case. It is true that a project remains in the attempt to discover a general rule distinguishing those cases where causal premises are crucial from those where they are not. But this question really concerns the justification$_1$ requirements; we are trying to discover when the subject must believe causal premises in order to be rational in believing a conclusion. The notion of a "crucial premise," then, although arising in the course of explicating justification$_2$, is actually *understood* by reference to justification$_1$. The latter is the more fundamental notion. (And hereafter it is what we will denote when we use the term without a subscript.)

But the externalist can try to press the point. After all, it is a hallmark of externalism that it stresses the necessity of non-epistemic conditions – appropriate causal grounding, for example, or production by a reliable

process – which often cannot plausibly be brought within the scope of internalist evidence. For the externalist this is acceptable; evidence (in the old internalist sense) plays, to say the least, a diminished role in his picture of human cognition. But for an epistemic internalist it is vital that there be no factor required for knowledge that lies beyond the reach of available evidence.

The answer to any attempt by the externalist to use the justification$_2$ condition as a stepping-stone to his epistemic theory lies in the close connection between justification$_2$ and knowledge. We must bear in mind that holding a non-epistemic theory of truth with respect to the T condition for knowledge does not prevent the internalist from developing a theory of knowledge. The non-epistemic factor represented by the T condition is clearly not "external" in any invidious sense: it does not lie beyond the reach of evidence. To complain that, for many beliefs, the evidence does not *guarantee* the conclusion would be to cavil; knowledge is justified true belief, not (necessarily) *guaranteed* true belief.

The considerations with respect to the non-epistemic aspect of the justification$_2$ condition in the Russellian analysis are perfectly parallel. The "external" factor is not inaccessible to evidence: it is simply the truth of beliefs for which, *ex hypothesi*, S has evidence, since the justification$_2$ condition includes the requirement that S be rational (justified$_1$) in the positive degree of credence he gives to the premises in question.[47] The fact that truth, as embodied in the justification$_2$ condition, is "external" should not give internalists any pause. For if *that* fact were taken to disqualify our analysis as a respectable internalist notion, then the complaint would prove too much: it would rule out any internalist theory of objective non-deductive inferential knowledge from the outset.

It is also quite clear that the T condition – whether a requirement for the conclusion or for crucial premises – is not at all similar to "reliable causal grounding," "production by proper function," or any of the other externalist requirements for positive epistemic status. To understand this point, it helps for us to think of the structure of the subject's reasoning for an inferred belief as a set of beliefs, connected in a particular way given the subject's basings, that forms an evidence tree existing in the subject's mind.[48] Once we think of the subject's argument in this way, we can see that to say that the conclusion is true or that some premise is true is not to say anything *new*, over and above what is found in the evidence tree itself.

But the claim that belief *p* was produced by an externally reliable causal mechanism, that belief *p* was produced by a particular module of a design plan and that this module was functioning properly, or anything of the sort, is new empirical information usually not found in the evidence tree itself. To defend the truth of a premise or a conclusion of his argument, the subject need use nothing more than the evidence tree he already has. But to defend claims about the reliability of the belief's cause or the nature of the cognitive module that produced it he will typically need empirical

information that the evidence tree, by itself, does not contain.[49] There is therefore no route from this solution to the Gettier problem to any general externalist condition on knowledge. An application of the term 'external' to a mere truth condition for premises would thus be a potentially confusing usage. It is perhaps for this reason that externalists have not been at all eager to agree that this *is* the correct solution to Gettier or to insist that acceptance of it is a victory for their position.[50]

This solution to the Gettier problem draws attention to another interesting fact: the request that we defend our beliefs against the charge that they are vitiated by Gettier problems is just the classic challenge of the traditional regress skeptic to defend our premises. In asking whether there is "some Gettier problem" with a given conclusion, a challenger is implicitly raising the possibility that some crucial assumption upon which we are basing our conclusion is actually false. If he specifies a particular Gettier problem that might beset our conclusion, the problem will, upon analysis, reveal the premise that is in question, and the correct response will be to show the reasons upon which we are basing that premise. Our interlocutor can construct a Gettier scenario in which some crucial premise further down in the evidence tree turns out to be false, and our response, again, will be to give the reasons underlying that premise, and so on, until we reach whatever terminus our theory of knowledge allows us to give to the questioning of the regress skeptic. Thus the Gettier problem brings us back to the Aristotelian regress that internalists have known about all along.

The Russellian solution, then, shows that the Gettier problem is not a new challenge that calls for externalist solutions but a very old challenge – to give reasons, and reasons for reasons – that calls for a very old solution: the discovery of truth in the course of our inferences. And for this purpose, internally accessible evidence has yet to be superseded.[51]

2 The Connection to Truth

A salient point of our internalist resolution of the Gettier problem is the centrality of the concept of rationality. But this very point is the target of one of the most common complaints against internalism – that in emphasizing rationality we have completely divorced epistemic rationality from knowledge.

Such a divorce is not unprecedented in the literature. Richard Foley, for example, argues that knowledge and epistemic rationality have so little to do with one another that one must simply choose which one to investigate at any given time. He therefore categorizes as "knowledge" cases where the subject has no reason whatsoever for holding a contingent belief (to the truth of which he has no direct access), but where the subject's inclination to believe is in fact reliably correlated with the truth of propositions of the kind in question.[1]

No internalist will be happy with such a proposal. But the externalist can press the question: does internalist justification have anything to do with truth? If not, then arguably it has no place in the analysis of knowledge. The J condition should have something to do with the T condition; if justification is to be an indispensable part of knowledge, we need to be able to defend its inclusion. Furthermore, as Stewart Cohen has argued, a non-trivial connection to truth seems required for distinctively *epistemic* (as opposed to pragmatic or ethical) justification.[2]

The problem of the connection to truth

The connection to truth criticism has been a stumbling-block for traditional notions of justification since before the emergence of the contemporary internalist/externalist controversy. Richard Fumerton argues persuasively that Bertrand Russell anticipated externalism in 1948 by his reaction to Keynes's theory of probability.[3] More recently, so staunch an internalist as Laurence BonJour cedes ground to externalism insofar as he worries that even the "third condition" for knowledge (that is, justification *aside from* a Gettier-blocking condition) may have an external component.[4] Russell, BonJour, and many others have been motivated by the intuition that

there must be a connection to truth in order for justification of *any* sort to obtain.

A survey of statements of this objection will not only illustrate the problem but also point towards the distinctions necessary to its solution. Fumerton, himself no externalist, states lucidly the motivation behind a rejection of internalist justification.

> The fundamental idea behind Goldman's reliabilism is straightforward enough. When a belief is justified it has a virtue. There is something good about it. From the epistemic perspective, virtue has to do with *truth*. The reason epistemologists want epistemically justified beliefs . . . is that having justified beliefs has *something* to do with having true beliefs. At the same time, we must understand justification in such a way that we allow the possibility of justified false belief. . . . The answer is to focus on the processes that produce beliefs.[5]

Alvin Goldman makes a similar point in his own words in a rather roundabout way. He first states that the best "doxastic-decision-procedure" (DDP) which should guide people in accepting beliefs will be one that produces an "optimal" combination of true belief and error avoidance. This, it seems to him, follows directly from the plausible idea that the goals of cognition are believing truth and avoiding error. He then argues that these goals present a problem for internalism.

> Unfortunately, the foregoing characterization of the right DDP ignores a crucial aspect of traditional epistemology. The foregoing conception rests on an "*externalist*" perspective: the perspective of a Godlike observer who, knowing all truths and falsehoods, can select the DDP that optimally conduces to true belief and error avoidance. Traditional epistemology has not adopted this externalist perspective. It has been predominantly *internalist* . . . . On the latter perspective, epistemology's job is to construct a doxastic principle or procedure *from the inside*, from our own individual vantage point. . . . The objective optimality of a DDP, on this view, does not make it right. A DDP counts as right only if it is "certifiable" *from within*.[6]

This internalist requirement that we certify any mode of arguing from within, says Goldman, cannot be satisfied. He therefore concludes that internalism is a "will-o-the-wisp" and that we should be satisfied with externalism, with its emphasis on the *de facto* optimal DDP.[7]

While Goldman stresses achievability, Bertrand Russell insists on what amounts to a form of externalism for more strictly analytical reasons.

> If induction is to serve the purposes which we expect it to serve in science, "probability" must be so interpreted that a probability

> statement asserts *a fact*; this requires that the kind of probability involved should be derivative from truth and falsehood, not an indefinable; and this, in turn, makes the finite-frequency interpretation more or less inevitable. . . . If the [inductive] principle is to serve its purpose, we must interpret "probable" as meaning "what in fact usually happens"; that is to say, we must interpret a probability as a frequency.[8]

While Russell does not use terminology exactly similar to that of the internalist-externalist controversy, the crucial concepts are here. Probability, according to Russell, must be "derivative from truth and falsehood." If it is related to truth and falsehood, it will not describe the rational credibility supported by a subject's evidence but rather will describe what in fact usually happens.

Michael Friedman also ties justification to truth in an externalist fashion when he addresses the possibility of an *a priori* defense of non-deductive inference forms in scientific methodology:

> The impossibility of such a justification follows, it seems to me, from two simple and fundamental facts: (i) there has to be some kind of link between justification and truth; a justification of scientific method must say something about its propensity to lead to truth; (ii) scientific method is not logically guaranteed of reaching true conclusions; it is an incurably nondeductive method. There is no inductive method that is more reliable in every logically possible world than every other method; consequently, there is no method that is *a priori* best, there is no method that is *a priori* the most reliable. We have to know facts about the actual world if we are to know which method is best; and we have to know facts about the actual world to know even that any given method has any chance at all of leading to truth.[9]

Friedman puts the point more clearly than does Russell. If there is "some kind of link" between justification and truth, then non-deductive justification arises only when one is using a "method" of induction that is reliable in the actual world.

David Miller brings the point down to the level of practical application, expressing impatience with those who emphasize good reasons.

> I often ask those who continue to find this . . . insight unacceptable the following question. *Which ticket would you prefer to draw in a sweepstake: the one bearing the favourite's name, or the one bearing the winner's?* The answer I receive is almost always that the winning ticket is, of course, the best one to draw; . . . But I am forcefully reminded after a semicolon's pause that, until the race is over, no one can know which ticket is the winning ticket, and as a matter of tactics the rational

> agent therefore prefers the ticket most likely to win ... It is the "therefore" here that takes my breath away. If the tactical preference for the most favoured ticket is not to be simply an underhand repudiation of the abstract preference for the winning ticket, then the agent must have conjectured that the ticket most likely to win actually will win.[10]

Miller concludes bracingly that "as far as rational thought is concerned, evaluation in terms of good reasons is a pure epiphenomenon"[11] – a comment that recalls strongly Goldman's reference to the external perspective of a Godlike observer. Both emphasize the desire to have beliefs that are in fact true, regardless of whether one possesses good reasons for them.

Finally, and most interestingly, Laurence BonJour has been so greatly influenced by the connection to truth consideration that he nearly capitulates to externalism on the quintessential internalist notion of having good reasons. BonJour puts the argument into the mouth of Alvin Plantinga, although it nowhere occurs in just these terms in Plantinga's own work. But BonJour is clearly worried by it himself.

> It seems apparent that an adequate third condition for knowledge must be one whose satisfaction yields an objectively good reason, not merely a subjectively good one, which is just to insist that justification or warrant must be objectively, and not merely subjectively, truth-conducive. The question then is whether ... oak tree experiences ... are objectively good reasons or merely subjectively good reasons for thinking that an oak tree is present. The answer to this question seems to hinge in large part on just how unusual this kind of case is in relation to the total class of actual and possible oak tree experiences. ... [I]f [this argument] is correct, then no purely internalist account of the third condition of knowledge will do, since some appeal to an external condition will be needed to guarantee the presence of an objective reason.[12]

Crucial distinctions

To evaluate the externalist argument that internalist justification is not connected to truth, we must make some essential distinctions. The first, frequently ignored in the philosophical literature, is the distinction between "sources of belief" or "belief-forming practices" and inference forms. In externalist writings one frequently encounters lists of belief-forming practices including everything from using induction to trusting our senses to trusting our friends. Alvin Plantinga groups together perception, the inclination to form beliefs about other minds, induction, memory, and the *sensus divinitatis*.[13] William Alston lists sense perception, memory, introspection,

and deductive and inductive reasoning all as "basic sources" of belief.[14] Goldman speaks of "doxastic decision procedures" as if all ways of deciding what to believe or of forming beliefs were simply processes to be compared for their effectiveness. But *prima facie*, there is an important difference between induction and deduction, on the one hand, and such "sources" as memory and perception on the other. The former are contentless forms of reasoning, while the latter are not forms of argument at all but rather types of experience that typically move us to form beliefs.

The importance of the distinction between inference forms and other "sources" of belief becomes more evident when we consider the distinction between an intrinsic and an extrinsic connection to truth, where the latter denotes some form of reliability in the world. In all of the quotations above where any determination can be made at all, it is clear that the authors are looking for an extrinsic connection to truth. Goldman speaks of the "objective optimality" of a DDP; Russell speaks of probability as a frequency; Friedman talks of a method's "propensity to lead to truth"; Miller is concerned with getting the "winning ticket" rather than getting the ticket one has the best reason to believe will win.

BonJour moves from a concern for an "objectively good reason" to a focus on the ratio of non-veridical oak tree experiences to veridical ones. But the concept of an "objectively good reason" requires careful analysis. In one of its meanings, "objectivity" is opposed to something like relativism. An objectively good reason would, on this interpretation, be one that is equally good for any reasoner who possesses it, as opposed to an argument that a reasoner merely happens to like or feels inclined to follow. Obviously, this is not what BonJour intends by an "objectively good reason." He means one that (in some sense) tends to yield true conclusions in the world outside of the subject's mind.

But there is reason to believe that the first concept of objectivity is the one that carries epistemic weight. We naturally want the epistemic resources to evaluate negatively the beliefs of lazy, biased, or even just confused individuals. We want to be able to say objectively that a belief is (or is not) rationally held by a given subject. As an internalist, BonJour should be especially open to the possibility that there is a discernible *intrinsic* virtue that arguments and non-inferred beliefs objectively either do or do not have. The possession of this virtue would mean that an argument constitutes an "objectively good reason" for its conclusion in the sense of a non-relativistically good reason, regardless of whether that argument form, or the inference from those sorts of sensations to that sort of conclusion, has in some further sense an external-world "propensity" to yield true conclusions.

What might such an intrinsic virtue look like, and in what non-trivial sense could we call it a "connection to truth"? The answer to this question will depend in part upon whether the belief in question is foundational. Foundational beliefs will be evaluated solely in terms of generative epistemic principles, while an evaluation of the epistemic status of inferred beliefs will

also require the use of transmissive epistemic principles. The epistemic principles will show that the foundational beliefs and the arguments based upon them have the relevant intrinsic virtue. It is therefore the epistemic principles themselves that must make evident the "connection to truth." There must be something the epistemic principles tell us about beliefs and arguments that shows those beliefs and arguments to be epistemically valuable.

For generative epistemic principles, not just anything will indicate an intrinsic connection to truth. The fact that, for example, a belief was produced by an externally reliable mechanism or by "proper function" would indicate only an extrinsic connection between the belief and truth. It is in no sense an intrinsic characteristic of a belief that it is produced by a properly functioning or externally reliable mechanism. Moderate foundationalists, with their emphasis upon merely probable foundational beliefs, are especially likely to endorse extrinsic foundational connections. Robert Audi, for example, writes that grounding in experience "seems to explain why a belief so grounded may be expected to be *true*; for experience seems to connect the beliefs [it] ground[s] to the reality constituting their object, in such a way that what is believed about that reality tends to be the case."[15] If this is simply an empirical statement that certain types of experiences seem to be (perhaps causally) related to external reality, then it is not strictly speaking an epistemological claim, and it requires an empirical argument. If, on the other hand, Audi means to give an account of the source of justification (or of any type of epistemically positive status) for experiential foundations, he is describing only an extrinsic connection to truth, since in a different world such beliefs might not "tend" to be true. Other moderate foundationalists have argued that some beliefs really are *intrinsically* probable.[16] This concept of probability, whatever its further analysis, does aim at describing an intrinsic connection to truth.

It would take a different book to argue for incorrigiblilist foundationalism of the sort we espouse.[17] But regardless of whether any sort of moderate foundationalism is defensible (and we believe that it is not), incorrigible foundations formed referentially clearly possess an intrinsic connection to truth, for whenever they are believed, they are by their very nature guaranteed to be true.[18]

It is more difficult to articulate what a connection to truth might mean for transmissive principles. As we saw at the beginning of this chapter, one philosopher after another has slid down the reliabilist slope to the conclusion that any form of reasoning that purports to confer probability upon a conclusion must produce a favorable proportion of true beliefs if consistently followed *as a practice*. If inference forms are separated from general "belief-forming practices," however, we can see that inference forms can have a connection to truth regardless of whether they are externally reliable.

There is a helpful parallel here between deductive and non-deductive principles. In his polemic against *a priori* justifications of scientific method, Michael Friedman emphasizes that scientific method is "not guaranteed of

producing true results" and is "incurably non-deductive." In contrasting deduction and induction, Friedman seems to imply that we know that deductive rules of inference are epistemically valuable because they are guaranteed to be reliable. The problem with non-deductive inferences on this view is that they possess no similar guarantee of external-world reliability.

It is true that when people start with true premises and reason correctly in accordance with the rules of deduction they invariably end up believing truths. But this does not mean that we need to look to empirical facts to investigate deduction's connection to truth. Indeed, the fact that correct deductions are defined by the rules of a deductive system enables us to investigate deduction non-empirically. The truth-preserving nature of deduction can be investigated *a priori*, both by our grasp of simple deductive rules and by way of proofs that display the consistency of a logical language in a way that can be clearly grasped.[19]

This fact about deductive logic points to the actual nature of its "connection to truth." Being transmissive, deduction merely preserves truth which must already be present in the premises. The truth-preserving nature of deduction is therefore the only "connection to truth" which it has in itself. Since its rules concern only connections between propositions, deduction does not possess a connection to external-world truth except insofar as it transmits the truth of external-world premises. But the fact that deduction *preserves* truth is not contingent but necessary. Hence, the distinctive deductive "connection to truth" is entirely a matter of the structure of deductive logic and owes nothing to the structure of the external world.

Correct non-deductive reasoning, like correct deductive reasoning, follows certain general and necessary rules.[20] The question of a connection to truth is interesting only in the case of reasoning that actually follows those rules, for we are no more interested in defending careless attempts at induction than in defending careless attempts at deductive reasoning. It is true that there is far more debate over the proper rules of non-deductive inference than over the parallel questions for deductive inference. But the correct rules of non-deductive reasoning – whether direct inference, Bayesian, or yet more complex forms – are neither equivalent to nor reducible to mere human "practices" or mechanisms generating readings in causal response to the environment, like watches or thermostats. The premises of non-deductive arguments confer a certain degree of epistemic probability on their conclusions, and this is not a matter of what the external world is like but rather a necessary relation that can be determined *a priori*. As with deductive logic, the connection of probabilities to truth is intrinsic.

Non-deductive inference and the epistemic interpretation of deduction

An obvious objection to any analogy between deductive and non-deductive inference is that the latter does not preserve truth in the sense that deductive

inference does. Deductive systems are checked by metatheoretic consistency proofs to demonstrate that they are truth-preserving; in deduction there is no partial credit for a system that *usually* yields true conclusions from true premises. But in non-deductive inference there is an ineliminable slippage between premises and conclusion, so that even a conclusion inferred from premises known with certainty may turn out to be false. Even if we say that non-deductive inference preserves or conveys probability, probability is not truth. It makes no sense to say that non-deductive inference preserves "probable truth," as if this were a species of truth, for truth does not come in degrees.[21]

What, then, is the point of referring to non-deductive inference as "partial entailment"? What does it mean to say that 'probable' means probably true? And how can the analogy between non-deductive rules of inference and the rules of deduction provide the former with an answer to the problem of the connection to truth?

We can start to answer these questions by considering the connection between deductive logic and probability at the level of pure (formal) semantics. Instead of assigning the values "T" and "F" to the propositional letters, we can assign "1" and "0." The results of applying deductive rules to premises with these values can then be seen as limiting cases of the results of applying non-deductive rules, since non-deductive rules will in general yield values intermediate between 1 and 0.[22] For example, we may assign the value 1 to the statements:

B: Nine out of ten Ss are Ps

and

P: *a* is an S

Here the principle of direct inference, which we shall explore in more detail in Chapter 7, ascribes a valuation of .9 to the conclusion:

C: *a* is a P

on the basis of B and P – or, to put it a bit more precisely, it evaluates the strength of the connection between B and P on the one hand and C on the other at .9. This is an entirely formal concept of partial entailment in which both the rules of non-deductive inference and the valuations are given no ordinary semantic interpretation (such as "true," "false," "certain" or "probable") and in which even the propositional letters are not thought of as "propositions."

Suppose, then, that we move to the level of informal semantics, where we do not merely assign valuations to propositional letters but actually give meaning to those valuations and think of the propositional letters as

representing meaningful propositions. At this level, deduction from premises given the values "T" or "F" admits of an entirely non-epistemic interpretation. Truth is non-epistemic; hence, we can define deductive entailments by speaking of which propositions must be true given the truth of other propositions. We need not make any reference to agents or epistemic categories such as *rationality* or *justification* in order to give an informal semantic interpretation for deduction from true premises. In contrast, non-deductive inference always introduces valuations other than T or F for propositions and thus requires (at the informal level) a semantics of probability.[23]

We notice, however, that deductive inference can be applied to premises labeled not "T" or "F," nor even "1" or "0" (which some might be tempted to treat as equivalent to truth or falsehood) but also to premises given numerical valuations between 1 and 0. For example, suppose that we assign a probability .9 to both '(P → Q)' and 'P.' The probability of Q need not, in the formal semantics of probability logic, be 1 even though Q is entailed by these two claims. What is preserved by deduction in such cases cannot, therefore, be simply the *truth* of the premises. So the problem of characterizing probability at the level of informal semantics arises already with the shift to a continuum of valuations between 0 and 1, even before distinctively non-deductive rules of inference are introduced.

What does deduction itself preserve under such a valuation? What do the valuations signify? An obvious answer is that, stretching the meaning of the term a bit, deduction "preserves" rational confidence about propositions. In the illustration above, for example, the probability of Q cannot be lower than .8.[24] This suggestion raises the possibility that at the level of informal semantics deduction itself is susceptible of an epistemic interpretation. The valuation "1" can be given the semantic interpretation "certainty" and the valuation "0" the interpretation "certainty of the negation of the proposition." When the premises of a deductive argument are all certain, a subject who reasons by correct deductive rules will give the conclusion the value 1. In these circumstances, it might seem that the use of such numbers is superfluous; yet when we provide a semantics for the numbers their relevance both to deduction with non-certain premises and to non-deductive inference becomes evident. The numerical values – or intervals, where appropriate – can be seen as credibility ascriptions representing the degree of confidence regarding premises or conclusions of a perfectly rational subject in a particular epistemic context. Certainty of the truth of a proposition and certainty of its falsehood are limiting cases of varying degrees of uncertainty. Non-deductive inference, like deductive inference, is able to convey probability, i.e. rational confidence regarding propositions.

John Maynard Keynes expressed a theory much like this, and in much the same terms:

> The terms *certain* and *probable* describe the various degrees of rational belief about a proposition which different amounts of knowledge

> authorise us to entertain. . . . [W]hile it is often convenient to speak of propositions as certain or probable, this expresses strictly a relationship in which they stand to a *corpus* of knowledge, actual or hypothetical, and not a characteristic of the propositions in themselves. . . . To this extent, therefore, probability may be called subjective. But in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. . . . When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is *rational* to entertain in given conditions, and not merely with the actual beliefs of particular individuals, which may or may not be rational.[25]

One advantage of such an epistemic interpretation of probability is that it makes evident the relation between probability and truth, since rational confidence is epistemically pertinent to the question of whether a proposition is true or false. Without specifying a particular "cut-off," we can say plausibly that a high rational confidence involves believing p, and that believing p means the same thing as accepting p as true or taking p to be true.[26] Rational credibilities help us to determine whether a subject may rationally believe, disbelieve, doubt, or remain uncommitted regarding a proposition. It is in this non-trivial sense that 'probable' means probably true: if a proposition is highly probable on the basis of one's evidence, it is rational on the basis of that evidence to accept the proposition as true. Inferential relations which convey rational confidence are therefore intrinsically *truth-directed*, i.e. aimed toward truth, since they tell us when and how inference permits a subject rationally to believe a proposition – that is, take it to be true. And this is the case even if the application of such inference forms is not truth-conducive, or successful, in the external world.

The asymmetry between inference forms and practices

The foregoing discussion highlights an asymmetry between inference forms and other "practices." It is true that inference forms can be *treated* as practices just in the sense that we might consider the human "practices" of following these inference forms correctly and might then investigate their successfulness in the real world. But to do so is to treat inference forms as belief-generating "black boxes" and to ignore the necessary connections of rational support between propositions. Other "practices" differ from inference forms in that one cannot reasonably claim that they express such relations. It is therefore not possible to reverse the process and to treat other "practices" as inference forms. The practices of trusting the senses, trusting friends, trusting memory, believing in the existence of other minds, and so forth cannot be *intrinsically* connected to truth as generators of foundational

beliefs, nor do they describe rule-governed inferential connections between propositions which purport to convey rational confidence in propositions by their very nature. For this reason, on our view, there will not be epistemic principles for these separate, specific practices. Rather, it will have to be possible to show an instance of trusting memory or believing in other minds to be rational in terms of principles describing types of intrinsically truth-connected foundational beliefs and inference forms (in this case, non-deductive inference forms) that intrinsically preserve rational confidence.

The fact that the claims for inference forms are different from the claims for general practices indicates a further asymmetry. Any attempt to investigate the reliability of a belief-generating mechanism or the success of a belief-producing practice necessarily requires the use of inference forms; but it is not true, conversely, that any attempt to investigate the affidavits of inference forms necessarily requires dependence upon the reliability of some other practice. Belief-forming practices such as trusting our friends or our senses have their only connection to truth externally; that is, if they are "connected to truth," it is only in the sense that they really do produce a high proportion of true beliefs. Therefore, their connection to truth can be investigated only empirically. But we wish to maintain the unfashionable position that, because inference forms purport to have an intrinsic connection to truth, it is possible to investigate them from a purely "armchair" or *a priori* perspective, without having to rely upon memory, the senses, testimony, and so forth.

The first part of this contrast is fairly obvious. Attempts to find out whether we usually get it right when we follow some practice are incurably empirical. To draw the conclusion that, by trusting perception, we form correct beliefs some favorable proportion of the time, we must gather evidence. And at a minimum, making the connection between evidence and conclusion requires using either deductive or non-deductive inference forms.[27]

A certain sort of Plantingian might respond that our beliefs that these practices are externally reliable are "properly basic." But without even commenting on the plausibility of such an assertion about proper basicality, we may note that holding beliefs about reliability in such a way would not involve *investigating* their truth, a process that is obviously inferential in nature.

It is, in any event, the higher claim for inference forms that is likely to prove more controversial. Alvin Plantinga frequently objects to any attempt to put the *a priori* on a different plane from other "sources" of belief, and he repeatedly quotes Thomas Reid to bolster the claim that all sources are in the same boat in that they must be taken more or less on trust.

> Why, sir, should I believe the faculty of reason more than that of perception? They came both out of the same shop, and were made by the same artist; and if he puts one piece of false ware into my hands, what should hinder him from putting another?[28]

One can, on Plantinga's view, take some sources on trust and use those to investigate the others, but it will be an arbitrary exercise to decide which sources to take as given. Hence, he concludes, there is no privileged "original position" from which to evaluate claims which (on his view) are "properly basic," be they the deliverances of the senses, of memory, or of reason.[29]

In arguing for such complete parity among all sources of belief, Plantinga relies heavily on the assumption that our justification for believing what purports to be an *a priori* truth is simply that a certain belief has "the peculiar feel that *a priori* beliefs have – that feel that somehow they just couldn't possibly be false. But of course," he continues, "Such a feel could be misleading."[30] We shall discuss further the "mere phenomenology" approach to the *a priori* in Chapter 5, but we will note here that Plantinga assumes without conspicuous argument that there can be no such experience as direct grasping of *a priori* truth. While he uses examples such as the Russell paradox to point out that people have been mistaken about what are (if anything is) *a priori* matters, he goes farther and implies that a subject's justification for believing even such a simple proposition as 2 + 1 = 3 is just a matter of "finding" himself believing it and "trusting" that the "feel" he has about it is not misleading in this particular case. Similarly, the best we can say, according to Plantinga, about the belief that the corresponding conditional of *modus ponens* is true is that we "just find ourselves with this powerful inclination to believe this proposition is true, and indeed couldn't be false."[31]

It is, however, important to distinguish those *a priori* truths which are so simple that we can hold them in our minds all at once and see their truth clearly and distinctly from those which, because they require a concatenation of steps, many of us cannot grasp all at once.[32] In the latter case, the proofs must be broken down into shorter steps. To say this is not, however, to concede that in *no* case are we able to perceive the truth of an *a priori* claim in such a way that we could not be mistaken about it. In those cases what we have is not simply a subjectively strong feeling about a proposition but a direct grasping of its truth. It is for this reason that no one who is capable of understanding statements such as 2 + 1 = 3 can be mistaken about them, whereas for most people, more complicated propositions of logic or mathematics cannot be seen with this self-verifying clarity.[33]

It may seem that in making this distinction, we are making a fatal admission. For inductive logic is a difficult subject, as is epistemology. Even if it is in principle an *a priori* matter that certain rules of inference connect propositions in such a way as to confer probability upon conclusions, must we not nevertheless "take on trust" our own ability to see this truth connection, since much of probability theory is too complex to be grasped all at once? Even if we can give an *a priori* argument that induction confers rational credibility, must we not rely upon memory, for example, to feel confident that we have not forgotten something or made some mistake when following the argument in its step-by-step form?

But the point we are making here is that it is not possible *in principle* to investigate external reliability without using inference forms, while it is possible *in principle* to see an inference form's intrinsic connection to truth without relying on "practices" that may or may not be reliable. Whether or not "we" have to rely on memory in considering a given putative law of deductive or non-deductive inference depends on who "we" are. Some subjects can hold more steps of an *a priori* argument in their minds at once than others. Some can see clearly and distinctly a conclusion too complex to be seen all at once by those less well-trained. But far from making the evaluation of inference forms into an empirical discipline, these differences among people of varying degrees of talent and mental power show that partial reliance on memory is really beside the point.

By their nature, *a priori* truths are the *sort* of thing that can be grasped in a self-evident fashion. Indeed, when contemplating 2 + 1 = 3 or the corresponding conditional for *modus ponens*, one has (*pace* Plantinga) exactly such an experience – *not* simply a vague feeling which might or might not be correct, but a genuine experience of seeing the truth of a proposition by reason of its conceptual structure. It is, then, merely a matter of extrapolation of our own powers and our own experiences of grasping the truth to imagine more complicated *a priori* matters similarly being understood in a clear and distinct fashion. It is a contingent fact that this or that person must break down the mathematical rationale for, say, Bayes's Theorem into several steps in order to grasp each part of the argument. Some other person may well be able to see it whole. The *in principle* nature of the subject matter is all that is required to distinguish inference forms from empirically based "practices." It is a necessary aspect of arguing for the reliability of practices that one use inference forms, but it is no necessary aspect of arguing for the truth-directed nature of inference forms that one trust some other, contingently reliable, faculty or practice.

Replies to objections

Some readers, and not only those of an externalist persuasion, may already have concluded that the notion of incorrigible *a priori* knowledge is a "will-o-the-wisp." For the present we wish to bracket that issue, though we will return to it in some detail in Chapter 5. But even laying that aside, we must confront some serious objections to the construal of epistemic principles as *a priori* and as analytic – true in virtue of meaning.

Despite the fact that he generally favors an armchair approach to epistemology, Richard Fumerton sounds a note of caution worth heeding. There is a danger that, in availing ourselves of the claim that epistemological propositions have *a priori* status, we will become dogmatic and classify as *a priori* epistemic principles whatever propositions we find convenient for evaluating the conclusions we wish to dub "justified." By this means, one could help oneself to principles guaranteed to deliver a non-skeptical conclusion.[34]

Since this is obviously an unacceptable use of the *a priori* appeal, how can *a priorists* avoid it? As Fumerton emphasizes, one important way for *a priorists* to keep themselves honest is not to be driven by the fear of skepticism. One must be willing to explicate principles independently of the conclusions one wishes to certify as justified and let the philosophical chips fall where they may.[35] It also "should go without saying," Fumerton points out, that in appealing to relations between propositions that one is able to "see" "one must be absolutely sure that one understands that about which one talks."[36] If there is supposed to be an *a priori* probability relation between propositions, one must really have a clear concept of it before appealing to it.

All of this is excellent advice, but there is more to be said. Fumerton takes the *a priorist* position regarding non-deductive inference to be that the probabilifying relationship between propositions is "unanalyzable."[37] He also asserts, in line with this assumption, that:

> the foundationalist who . . . seeks to avoid both epistemic and conceptual regress concerning justified beliefs about probabilistic connections by embracing a Keynesian conception of epistemic probability, will refuse to offer an argument for the various epistemic principles . . . he endorses.[38]

At this point worries about dogmatism understandably arise. We discuss the nature of argument in *a priori* contexts more fully in Chapter 6 in the context of the defense of deduction, contending that argument in *a priori* contexts is really explication. Given that this is the case, one can be both a foundationalist and an *a priorist* and hold that, for many epistemic principles, it is possible to give explicatory "arguments" in the form of conceptual analysis, although there will be some relations of ideas so simple that they do not admit of further analysis.

An indispensable requirement for the honest *a priorist* is that he analyze as far as he can, both to achieve entirely clear understanding for himself and to answer the objections of a skeptical audience. By way of conceptual analysis the *a priorist* should aim to present a convincing case for epistemic principles that have as wide an applicability as possible. For example, some version or cousin of Ockham's razor may very well be an *a priori* epistemic principle. It is nevertheless quite unsatisfactory for an epistemologist to say only that simple theories are more likely to be true than complicated ones. Even the addition of a *ceteris paribus* clause will scarcely suffice to render this bare assertion an example of *a priori* epistemological reasoning worth the name. What is the meaning of "comparative simplicity"? Can simplicity be quantified? Why are simpler theories more likely to be true than more complex ones? Can we give a rigorous proof of this claim using still more fundamental theorems of probability? All of these questions may have entirely adequate answers, but part of the epistemologist's job is to try to

answer them, even for the sake of his own clear vision of the principle he is claiming to know *a priori*. There will come a point where further analysis is impossible and where clear conceptual connections must be grasped directly, but probability theory is difficult enough that such an end of analysis will in many cases be a long time in coming. With this demanding version of an *a priori* approach on the table, the skeptic need not fear that he will receive simplistic answers based on an *ad hoc* attempt to guarantee common sense results.

The discussion so far will make it fairly clear that we take *a priori* epistemic principles to be analytic. Fumerton takes the only plausible form of *a priorism* to be rationalist, i.e. the position that epistemic principles are synthetic *a priori*. Not only does he consider the probability of sheer dogmatism (without any genuine *a priori* grasping) to be especially high when philosophers take epistemic principles to be analytic, he also believes that an analytic approach will trivialize the discussion.

> It seems to me a little difficult to suppose that the many skeptics and those who took them seriously were all simply misusing language. However implausible we might view skepticism about the physical world, are we really to maintain that such skeptics were simply contradicting themselves? Can we really dismiss the skeptical challenge by exclaiming that we just *understand* rationality in such a way that it follows from the concept alone that sensations make it rational to believe propositions about the physical world? In short, the solution seems too easy.[39]

This objection is echoed by Michael Friedman, who expresses exasperation at:

> attempts to justify induction by appealing to the *meaning* of 'justified' or of 'rational' or of 'inference' . . . . It is not very comforting to be told that scientific method is justified or rational in virtue of the meaning of 'justified' or the meaning of 'rational' . . . .[40]

Though we are defending an analytic version of the *a priori* approach that Friedman finds "uncomforting," his and Fumerton's impatience at an attempt to justify induction by linguistic fiat seems reasonable. A successful *a priori* defense of inference forms will have to do better than just smuggling 'rational' in as part of a stipulative definition of a form of reasoning. The epistemologist will have to show that certain forms of inference are intrinsically truth-directed, that it is rational to reason in accordance with them. For induction, this will require explicatory argument in terms of concepts such as reference class, sample, and randomness (see Chapter 7). For inference to the best explanation, the *a priorist* will need to parse out concepts such as simplicity and explanatory power, and for Bayesian

inference, concepts such as comparative likelihood and conditional independence. There is much important work to be done in determining whether IBE is a primitive form of inference or can always be modeled in Bayesian form, and this, too, will require careful attention to the concepts of explanation and probability and to the ways in which Bayesian probability theory might serve to model various theoretical virtues promoted by the advocate of IBE – the "loveliness" of a theory.[41] The emphasis, then, is on concepts and their connections, with words playing a role only insofar as they are used to designate concepts.

In point of fact, the method of conceptual analysis as we conceive it is less prone to produce unargued statements of "insight" than a position that depends upon the synthetic *a priori*. The advocate of analyticity bears the burden of showing that his positive evaluation of a form of inference really does bring together detailed, reflective understandings of basic concepts of rationality. While the concepts involved will ultimately be his own concepts, as he can have only defeasible access to the concepts of others, his position will have plausibility for others (and, if he is honest, for himself) only if it can be defended by careful analysis. The advocate of the synthetic *a priori*, on the other hand, has no such burden, since he does not claim to be analyzing anything at all.[42]

What of Fumerton's concerns regarding the skeptic? Does analytic *a priorism* imply that the skeptic's concerns are trivial and easily answered? Has the skeptic simply been "misusing language" by declaring ordinary beliefs unjustified if, in fact, they are justified as a matter of analytic, *a priori* truth? None of this follows if we work with a notion of analyticity that focuses on concepts rather than language. *If* the skeptic has been wrong, he has in fact been conceptually and therefore analytically wrong on our view (given certain assumptions about the skeptic's own concepts). But this implies no derogation to the skeptic. The concepts involved in probability theory are complex and difficult to grasp with complete clarity. Even in deductive logic there are proofs so complex that people make mistakes about them. In mathematics, a provable theorem is declared to be "trivial," but it is only when 'trivial' is thus used as a term of art that we will be able to declare the skeptic's concerns "trivial" – and that, only after we have answered him decisively.

This version of analytic *a priorism* also avoids the danger of fragmentation, a charge leveled at Roderick Chisholm and John Pollock by Stewart Cohen. Cohen points out that their principles "are nowhere united by a general theory that explains why those beliefs are justified under those conditions. In a sense, [their] theories do not tell us what justification is."[43] Fumerton raises the same issue with specific reference to analytic versions of *a priorism*.

> One is sorely tempted to suppose that philosophers who take epistemic principles to be analytic do normative epistemology by simply listing their prephilosophical beliefs, deciding what they do infer the

> propositions believed from, and proclaiming the epistemic principle sanctioning such inferences to be analytic. But what exactly do all these inferences have *in common* that makes it plausible to claim that they fall under a single concept of rational inference? Is the concept of probabilistic or evidential connection simply a disjunctive concept? . . . EITHER *p* is a proposition about the physical world inferred from truths about sensation OR *p* is a proposition about the past inferred from memory states OR *p* is a proposition about the future correlation of properties inferred from past correlations of properties . . . . It surely makes sense to ask: "In virtue of what do both sensations and memory experiences make probable, respectively, propositions about the physical world and propositions about the past?"[44]

It does, indeed. But the association of such fragmentation with analytic *a priorism* arises from the idea that the analytic *a priorist* will take, say, the relation between memory-like experiences and beliefs about the past to be a fundamental evidential relation.[45] But it is by no means clear that there is such a thing as a "memory-experience inference form." A far more plausible version of the position would take (for example) Bayesian reasoning to be an inference form and would then try to show how reasoning from memory-type experiences to beliefs about the past falls under the category of an inference to the best explanation, understood in Bayesian terms. Similarly, induction could be taken to be a genuine inference form, and inferences from the past to the future, from sampled balls in bags to unsampled ones, and so forth, would all be *cases in point* of the inductive inference form. This approach might still end up listing more than one unique justificatory inference type, but the number of types would be far smaller than Fumerton envisages.

Perhaps more importantly, discussions of Bayes's Theorem or of the proportional syllogism (see Chapter 7) do not involve reference to some highly specific, concrete form of experience that is taken to justify a specific conclusion. Rather, these are contentless forms of reasoning, and it is up to the epistemologist who wishes to defend inferences having specific content (inferences from past experiences to future expectations, from sensations to beliefs about a mind-independent physical world, and so forth) to show how they fall under these categories. The inference forms themselves – deductive or non-deductive – are united (or so one will try to show by conceptual analysis) in that they all convey rational confidence by entailment or partial entailment, as discussed above.

Friedman's complaint that *a priori* approaches are not "comforting," however, leads us back to the connection to truth issue. Those who have embraced the externalist version of a truth connection requirement may simply feel that an intrinsic truth connection is not good enough. Suppose that one is concerned about having true beliefs about the world outside one's own mind. Incorrigible foundationalism of the sort we advocate will

not grant foundational status to truths about the external world. And even if there were external-world beliefs that could be known foundationally, pure deduction will not get us very far in, say, scientific investigation. Yet non-deductive inference forms do not preserve truth perfectly. They can be proven to convey a particular degree of rational confidence, but one could reason in accordance with intrinsically impeccable non-deductive inference forms from intrinsically truth-connected premises and still end up believing a set composed entirely of falsehoods. It would require some ingenuity to dream up such a situation, but surely a sufficiently diabolical Deceiver can be found to do the trick.

Many philosophers are inclined to feel that it is unsatisfying to bestow the term 'justified' upon beliefs produced by such a process. They want, as Bertrand Russell says, to think of probability as "what in fact usually happens," of justification in terms of success. Or, if the rational inference "practices" we follow do not *actually* produce a high proportion of true beliefs, perhaps they should at least have some sort of counterfactual propensity to produce true beliefs in a set of possible worlds. Otherwise, so the argument goes, why should we care about mere rationality at all? On this argument, it seems that an intrinsic connection to truth is nice as far as it goes, but only an extrinsic connection has any "cash value."

Perhaps nothing said here will move those who have such an intuition. There are, however, two relevant responses. First, Fumerton has argued convincingly that any sophisticated version of externalism will be open to a similar objection.[46] He points out that a belief-forming process type might operate only once in the actual world and might quite accidentally give a true conclusion in that instance. Since "accidental justification" bothers externalists a good deal (as witness their response to the Gettier problem), the sophisticated externalist will want to invoke counterfactuals or propensities in order to indicate a genuinely *nomological* connection between the operation of the belief-forming mechanism and truth. But once the externalist has shifted his analysis in a counterfactual direction, it is logically possible that all of our own *actual* inferences are just those in which the process type does not manifest its propensity to produce true beliefs. In fact, it might happen that *no one* in the actual world arrived at a true belief by means of such a process type. It is logically possible that a fair die should always land with the six side up; similarly, it is logically possible that a process that is "reliable" in terms of the set of possible worlds should, by sheer bad luck, always produce false beliefs in the real world. But in that case, the externalist would be calling beliefs "justified" although they were produced by a process type that was *de facto* unreliable (even if reliable counterfactually or nomologically). He would thus lose his supposed advantage over the internalist: what "cash value" would the "reliable" process have in *our* world?

Second, the internalist can challenge directly the implication that he is insufficiently concerned about truth – even truth about the external world.

The most die-hard internalist will agree that it would be a good and desirable thing for our inference forms to produce a hefty proportion of true beliefs in the real world. He may even believe that the world is stable and orderly, that no clever Deceiver exists who stirs up reality so as to defeat our inductions, and that therefore induction (for example) *does* usually produce true beliefs. But if so, he believes this as a result of an inference. No mundane subject has direct, infallible access to the nature of the physical world. Even Plantinga's "properly basic beliefs" are "properly basic" only because a particular subject does not happen to infer them and because they are produced reliably by some non-belief-dependent, truth-conducive module of the subject's design plan. They are not the result of an infallible insight about the external world.

But as we have already argued in describing the asymmetry between inference forms and practices, anyone who wishes to *investigate* claims about the stability of nature or the reliability of this or that process (rather than making such claims without argument and dubbing them "properly basic") will have to gather evidence and draw conclusions from that evidence. The argument for those conclusions will itself have a structure that either is or is not intrinsically truth-directed.[47] If someone decides that nature is uniform, for example, and if he then ascends to the level of epistemology to investigate the affidavits of his inference form, there is no reason for him to feign ignorance by treating the inference as a structureless "mechanism" and engaging in a purely empirical investigation as to its reliability.[48] It is surely worth his while, if he has the capability, to look into the question of whether his inference form has the desirable quality – and the quality it purports to have, in virtue of its premises and the inferential rules it follows – of making its conclusion rationally believable. It is worth his while precisely because he is interested in the question of whether his object-level conclusion about the stability of the world is *true*. Therefore, a concern about the truth-conduciveness of our practices leads directly to gathering evidence and making an empirical argument that (for example) the world is a relatively stable place. And a concern about the structure of that argument leads to an investigation of intrinsic truth connections.

Far from abandoning a concern for truth, then, internalists have taken the most direct route to discerning it, and to discerning that they have discerned it. The externalist approach offers only one empirical investigation after another, the final vindication of our beliefs ever receding before us, our hopeful spirits buoyed by promissory notes to the effect that we are justified *if* the process we have just used is reliable. And by conflating inference forms with empirical practices, the externalist ignores the special claims of the former, not bothering even to investigate the possibility of an intrinsic truth connection. It behooves us, precisely if and because we care about truth, to make that investigation.

3 Internalism, Externalism, and the Metaregress

Externalism and internalism: A first approximation

The burden of the previous two chapters has been primarily defensive: we have been arguing that two key challenges leveled against traditional epistemologists by self-described externalists are not actually damaging. But now we shall be examining externalism itself, and for this purpose it is no longer sufficient to characterize externalism by pointing to its adherents. To avoid ambiguity and needless wrangling, we now stipulate a partial definition: externalism requires that, for a belief to have positive epistemic status (whether one calls that "warrant," "justification," or simply "the characteristic that distinguishes mere true belief from knowledge"), there must be a reliable connection – typically, although not necessarily, nomological or causal – between the subject's believing that *p* and the truth of *p*. Such a connection may be expressed by saying that one's belief tracks the truth that *p*, that the belief possesses a feature F that is a reliable indicator of the truth of *p*, or that it is generated by a source M which is a reliable mechanism for the production of true beliefs.[1] Furthermore, internal rationality is neither necessary nor in general sufficient to confer upon S's belief that *p* any positive epistemic status.

This definition is not much use until the concept of internal rationality is spelled out, and we will be developing key aspects of this concept at more length later. But at a first approximation we can say that the internal epistemic status of S's belief that *p* – and in particular the degree of credence that a rational subject would give to the belief, given the evidence (if any) on which S grounds his belief – must be evident to a subject who has complete access to S's mental states and who has a completely clear grasp of the epistemic concepts relevant to the belief in question. And a belief has internal positive epistemic status only if, in virtue of that internal credibility, it is rational for S to hold. Only beliefs that meet this standard are internally rational.

On this understanding of internal rationality, contingent propositions are relevant to a belief's internal epistemic status only if either they express empirical information relevant to an inferred conclusion or else they are

truths about the subject's own internal states (e.g. the claim that he believes that *p* or that he is indeed basing his belief that *p* on another belief that *q*). This point makes the internalist nature of the rationality in question particularly evident, for it means that the subject *himself* could grasp the epistemic status of his own beliefs in a decisive way – that is to say, in a way that did not simply generate another reasonable, empirical question about the epistemic status of the belief about epistemic status – if only he were given a sufficient degree of conceptual sophistication.[2] Of course, many knowers are conceptually unsophisticated and perhaps most are typically unreflective. But if S *were* to acquire the relevant conceptual resources, he would be able to determine the epistemic standing of his own beliefs by pure introspection, since he has (so we would argue) direct access to his own mental states. Such conceptual sophistication and introspective clarity are not themselves necessary conditions for internal rationality. But when these conditions are fulfilled, internal rationality suffices for what Alston has called "fully reflective justification."[3] And this is so because the standards of rationality are necessary truths knowable *a priori*.

A note about basing is in order here. Internalists who maintain the epistemic transparency of the basing relation need not take a position on the analysis of that relation; but if they do, the analysis will be subject to that constraint. If the relation is partly or wholly causal, as several authors have urged,[4] it must be a special kind of causal relation the holding of which is introspectable.[5] Basing may, on the other hand, be *sui generis*, not reducible to any other kind of relation, an alternative little explored in the literature.[6] In any event, nothing in the present argument turns on any more specific construal of the basing relation than that it satisfy the introspectability constraint.

The sort of externalist we are concerned with here finds a strong connection between internal rationality and positive epistemic status to be unacceptable in two ways. First, he will be willing to grant his honorific epistemic term to beliefs that do not meet this standard of rationality, beliefs that even a subject with full conceptual understanding could not evaluate decisively. So, for example, Alvin Plantinga defines "warrant" in terms of the process that produces the belief in question; his definition would allow in theory *any* sort of belief to be "properly basic" if it were produced by the correct module (*inter alia*, a truth-directed module) of the correct sort of design plan (*inter alia*, a good, i.e. reliable, plan) operating in the correct sort of environment. The fact that "properly basic" beliefs are neither identified by their intrinsic nature nor confined to those types of beliefs that meet internally discernible standards makes it quite clear that a belief may be properly basic on Plantinga's view without being internally rational.[7] Even for inferred beliefs, Plantinga explicitly endorses the possibility that beliefs formed in accordance with an anti-inductive principle might be "rational" for some creatures, not in virtue of evidence available to them (e.g. that a particular situation is chaotic) but solely in virtue of the

nature of their design plan.[8] He stresses that inferential "rationality," in his system, is entirely a contingent matter; therefore, one cannot evaluate the Plantingian "rationality" of an inferred belief without knowing facts about the design plan in question. Hence, inferred beliefs as well do not require internal rationality in order to have warrant.

Second, and perhaps more interestingly, even a belief that meets stringent internal requirements for justification – e.g. a belief inferred by an *a priori* certifiably rational inference form from (let us say) infallibly known foundations would *lack* externalist positive epistemic status if it lacked the externalist's required connection to truth. For example, if S arrives at his belief in some contingent non-evident proposition *p* by using any non-deductive inference form, it is logically possible that S's conclusion is false even if all of his premises are true. Yet he could be internally rational in holding the conclusion. Further, it is *possible* (as we acknowledged in the last chapter) that the "practice" of using such an inference form is not reliable on either an "actual world" or a "possible worlds" model of reliability. Internal rationality is therefore not in general sufficient for externalist positive epistemic status.[9]

This is not to deny that there may be considerable extensional overlap between the sets of beliefs that have internalist justification and those that have externalist positive epistemic status. It might be true *in fact* that the belief-producing processes the internalist would endorse, including the following of non-deductive inference forms, would produce the "right" proportion of truths to meet an externalist's requirement. Furthermore, one could view infallible "processes" that meet internalist requirements – e.g. direct acquaintance, deduction (held in the mind all at once) from infallibly known premises – as "super-reliable"; and so, if these "processes" met any other external requirements of a particular version of externalism, they might also confer externalist positive status. So internal rationality may on particular occasions suffice for externalist positive epistemic status, but it is not *in general* sufficient for externalist positive epistemic status.

The version of internalism we advocate holds that internal rationality is both a necessary and a sufficient condition for any positive epistemic status. This is, in the terminology of Chapter 1, justification$_1$ rather than justification$_2$. On our account neither 'justification$_2$' nor 'knowledge' is a term of pure epistemic appraisal. The claim that S's belief that *p* is justified$_2$, like the claim that he knows that *p*, entails that his belief is justified$_1$ and hence that it has positive epistemic status. But each of these claims goes beyond that entailment to include truth, either of the crucial premises on which S bases *p* or of *p* itself; and truth is a non-epistemic property.

Between this strong position and externalism as we have defined it there is logical space for hybrid positions that hold internal rationality to be necessary but not sufficient, or sufficient but not necessary, for positive epistemic status. In our view such hybrids are apt to be vulnerable to some of the same problems as pure externalism; but since such views have little

currency and, we think, little intuitive appeal, we will largely ignore hybrids and concentrate on the pure position.

Object level and metalevel

Suppose that a subject believes p, "The sun will rise tomorrow." He believes this, let us say, on the basis of numerous premises q_1, q_2, . . . , among them q_1, "The sun has risen on all of the mornings I can remember in my entire life." He is inferring his conclusion, *inter alia*, by way of an inductive extrapolation from q_1 to p. One of the constructive uses of skepticism is to press the question from p back to q_1, q_2, . . . , from each of these back to its premises, and so on, until one reaches epistemic bedrock. This is the classical regress argument due to Aristotle.[10] The terms in this regress are p and all of the subject's supporting reasons both for p and for its premises. We shall refer to the tree-like structure of these as the *Aristotelian* or *object level* regress for our subject S's belief that p, and we think of it as spreading out "horizontally" at the object level.

In this particular regress no terms of epistemic appraisal appear. This is a consequence of the fact that p is not a belief that employs epistemic concepts and S's reasons for believing that p do not themselves employ epistemic concepts – surely a realistic assumption for most ordinary believers. Yet several natural questions arise about p, questions in which epistemic terms are employed. Is S justified, at t, in believing that p? Were someone to believe that S is justified in his belief that p, would that belief itself be justified?

These questions are not answered at the object level, although information in the object level tree is obviously relevant to their answers. When one makes a statement concerning the epistemic status of S's belief, such as "S is justified in believing p at t" or (if one is a Plantingian) "S is warranted in believing p at t," one is ascending to the next level "up," as it were, with respect to p. We shall call the level at which Jp and any arguments for Jp would appear the *metalevel* for S's belief that p.

The set of metalevel statements includes not only the conclusive claim that S's belief has a particular epistemic status but also statements relevant to it – for example, "S believes that p on the basis of q_1 by way of inductive extrapolation," or, even more interestingly, "An inductive extrapolation based on premises of such-and-such a sort confers such-and-such a degree of rational credibility upon its conclusion." As we shall discuss below, what one considers to be a metalevel claim will depend upon one's epistemic theory, since different theories give different analyses of the very notion of justification, and some theories import concepts like warrant that have their own set of requirements. An externalist might use a claim about the reliability of induction as part of a defense of the claim that S is externally justified in believing p. Yet if any proposition *is* necessary for the defense of claims about the epistemic status of p for S, it belongs at the metalevel for p.

If we are discussing not *p* but *Jp* (where 'J' is understood to be relativized to S at t), we have moved to the metalevel for *p*. But the question of whether someone is justified in believing *Jp* moves us up yet again. The proposition *JJp* would appear only at the next higher level, the metalevel with respect to *Jp*.

In what sense do such levels "exist," if indeed they do exist? If S believes that *p* but no one, not even S, has any beliefs about the epistemic status of *p*, should we say that there is no metalevel for S's belief that *p*? Certainly many subjects hold beliefs without having any *clear* metalevel beliefs about them. Plausibly many subjects hold object level beliefs without holding any metalevel beliefs about them whatsoever, even tacitly. And it may be that no *other* person holds metalevel beliefs about S's beliefs, either.

Nonetheless, some metalevel facts follow from the nature of the object level itself. If S believes *p* in the way represented by a particular object level structure (e.g. non-inferentially or supported by a particular evidence tree, using a particular set of basing relations), his belief that *p* has a particular degree of internalist rational credibility in virtue of his cognitive structure at the object level, regardless of whether he or anyone else actually holds beliefs about *p*'s epistemic status. Furthermore, S's belief arises in some way or other, and although its externalist status will often not be determined solely by the nature of the object level, still either it does or it does not satisfy some set of externalist requirements. In either case, the *existence* of an object level for *p* entails the existence of at least one belief that *has* some epistemic status – justified or unjustified, warranted or unwarranted, and so forth.

In speaking of the metalevel for S's belief that *p*, then, we mean to speak of the level at which both claims ascribing epistemic status to S's belief that *p* and arguments defending such claims, if any are required, would appear. And by extension we will use this terminology to represent what *could be said* by an epistemically knowledgeable person – that is, a person who had full understanding of the relevant epistemic concepts and who had full access to S's mind – about the epistemic status of *p* for S at t.

Internalist metalevel vs. externalist metalevel

The fact that a belief meets the requirements of internal rationality need not entail that it meets some external criterion; if it meets the latter, that may be a contingent fact not entailed by anything internal. S's belief that *p* might, for example, be the conclusion to a non-deductive argument at its own object level. On some externalist models, a belief might fail to meet external requirements even if it were known *a priori*. Plantinga is particularly insistent that the *a priori* has no special status but is just another belief source on a par with others. Hence, if someone *knew a priori* that a type of reasoning was rational to follow, or even that a mode of access rendered a set of beliefs infallibly known, on Plantinga's model that epistemic belief *itself*

would not automatically satisfy externalist requirements for warrant. If any such belief does have externalist positive epistemic status, it must have a metalevel that differs crucially from the metalevel required by our version of internalism – a direct consequence of the fact that in the case of *p* as described, internal rationality is not automatically sufficient for externalist positive epistemic status.

Given externalism as we have defined it, a defense of the metalevel claim, "*p* is warranted" or "*p* is justified in an externalist sense" will have to contain a contingent statement to the effect that *p* meets the relevant reliable truth connection requirement. We will refer to the externalist's truth-connection(s) as "R property (or properties)" and to the relevant statement that must appear at the externalist metalevel as the "R-property statement." The R-property statement takes the form

> X: Beliefs produced by process (source, mechanism) M or having the feature F satisfy the R requirement (have the property R, the reliable connection to truth).

The internalist metalevel, in contrast, need never have a statement like X, for internalists need not concur in the requirement that beliefs with positive epistemic status have an R property. This point is important and must be stated clearly. The internalist need not concede that the truth of a statement like X is required for justification.

This point is not widely appreciated by externalists, who often assume that *even internalists* must or should concede the importance of reliability. William Alston, for example, argues that internalists must acknowledge the overarching importance of reliable belief formation because of the need for a connection to truth.

> Even those, like Roderick Chisholm and Richard Foley, who deny that likelihood of truth or reliability of source is required for being justified in believing that *p*, still recognize it as a constraint on the concept of epistemic justification that a belief's being justified is a good thing *vis-à-vis* the aim of believing what is true and avoiding what is false. And if that is the central goal by reference to which candidates for the status of justification are to be measured, then the notion of reliable belief formation must also be of central importance to epistemic assessment. Whatever one's epistemological theory, one can hardly avoid recognizing the crucial importance of the reliability of belief formation.[11]

Michael Bergmann refers to a "trustworthy source requirement" in this same fashion and implies that internalists who have agreed that beliefs must be "likely to be true" in order to be justified have thereby acquiesced in a trustworthy source requirement.[12]

We have already addressed at length the connection to truth issue, but it is interesting to see how stubborn is the notion that everyone, internalists included, must acknowledge the need for a general reliability requirement of some sort for positive epistemic status. It may be true, as we have already discussed, that a particular set of sufficient conditions for internalist justification – for example, being infallibly, non-inferentially known – also satisfies a very broadly understood version of a trustworthy source requirement. But the internalist should staunchly insist that coming from a trustworthy, i.e. reliable, source is neither necessary nor in general sufficient for justification. It is not necessary, for an internalist can take a belief to be justified if its source does not in any sense tend to produce true beliefs. It is not sufficient because all of the external conditions one could add, even the elaborate system of design plan, environment, and so forth required for Plantingian warrant, neither constitute nor entail that a belief has internalist rationality.

Plantinga drives this home with his example of the anti-inductive Alpha Centaurians.

> [I]t might be statistically likely that an Alpha Centaurian who is such that nine out of ten As he has come across are Bs, is also such that the next A he comes across will not be a B. Perhaps these Alpha Centaurians are furnished by a benevolent creator with inclinations to form, in such contexts, the belief that the next A will not be a B; and perhaps this inclination leads them to form true beliefs on most occasions.[13]

Under such conditions, Plantinga concludes, an inductive argument that would make its conclusion "epistemically probable" for us would not make its conclusion epistemically probable for the Alpha Centaurians. His point in no way turns on internal conditions such as the Alpha Centaurians' access to a track record for such anti-inductive policies. And Plantinga insists that, under the conditions, they could be *rational* creatures. It would be difficult to find a clearer example in the externalist literature of the vast gulf between internalist and externalist conceptions of rationality, between intrinsic and extrinsic connections to truth.

Since the internalist should deny any general requirement for reliability, he should deny that the defense of an ascription of positive epistemic status requires a statement like X – should deny that it must appear at the metalevel. But the externalist, obviously, cannot do without it. Except where an externalist would concede that the satisfaction of internalist requirements entails the satisfaction of externalist requirements, the externalist requires an *extra condition* for positive epistemic status – the presence of an extrinsic connection to truth – which has no parallel at the internalist metalevel.

Both internalists and externalists, it is true, will require some statements of the form, "S believes that *p*," "S has access of such-and-such a sort to *p*,"

and so forth, and these are contingent statements, since even the existence of S is contingent. But the truth of such statements, according to access internalists, is knowable by the subject himself by acquaintance with his own mental states, and hence their presence at the metalevel is perfectly compatible with the requirements of internalist rationality. More important for our point here, however, is the fact that contingent R-property statements are *not* knowable without additional empirical information since they are not about the subject's object-level beliefs *per se*. One can exhaust all parallel statements about the subject that would appear at both the externalist and the internalist metalevel; the R-property statement remains an irreducible extra for the defense of positive epistemic status in the externalist metalevel.

We refer to a "defense" of the claim that *p* has externalist positive epistemic status in order to forestall a peculiar form of dogmatism to which some externalists might resort. In the Plantingian system, a great many things are taken to be "properly basic." Indeed, it is quite impossible to say ahead of time that some belief-type *could not* be properly basic in Plantinga's terms.[14] Given that this is the case, a Plantingian might say that *p* is properly basic and that *Wp* is properly basic and that *WWp* is properly basic, and so forth, *ad infinitum*, arguing from this that a Plantingian metalevel for *p* does not have to have any statement such as X. In fact, on such a view, the metalevel for *p* might contain just *Wp* and no other statements whatsoever about S's belief that *p*.

But even if the externalist is prepared to go to such desperate lengths, two points remain salient. First, the truth conditions for the statement *Wp* are, in the case of *p* as stipulated, different from the truth conditions for internalist justification. That *p* has the relevant extrinsic connection to truth is part of what *Wp* means. In this sense, X is implicitly present in *Wp* even if the externalist coyly refuses to unpack it and state X as a separate claim. Second, simply to say "Wp" is not to offer any sort of defense of the positive epistemic status of *p*. And it is evident that, if X is one of the truth conditions for *Wp*, *Wp* does not satisfy any criterion for being self-evident. It is certainly a statement that is open to question, a statement for which one could reasonably request a defense, and simply to create a hierarchy of unargued (and even unexplained) statements of the form *Wp*, *WWp*, etc., is a form of dogmatism that it would be uncharitable to make definitional of externalism.

The fact that a statement like X is present at the externalist metalevel and absent at the internalist metalevel draws attention to a fundamental asymmetry between the two positions. The internalist metalevel requires *no new empirical information*. Reliability claims need not, and *should* not, be regarded by internalists as metalevel in any sense at all. If *p* is inferred, the beliefs in the object level structure of reasons will stand in a certain relationship to one another, the foundations will either have or not have the intrinsic truth connection, and these aspects of the object level tree will

determine the rationality or irrationality of an inferred belief. As for the claims that S does hold the belief in question and that he is basing his belief on such-and-such reasons, these are not premises of the object level argument, but they can be "read off" of the object level argument as it exists in S's mind. That is to say, each of them has a counterpart in a feature at the object level, and in that sense their appearance at the metalevel is not the introduction of anything new. What will not appear at the metalevel are additional empirical statements, such as X, about factors that have no counterpart at the object level.

This asymmetry is not built into the standard definition of internalism. Nevertheless, we believe that a consistent access or mental state internalist ought to adopt the position sketched here with respect to metalevel information; and if he does, then as we mentioned above it will follow that a subject can, in principle, determine the epistemic status of his beliefs from his armchair. Garden variety internalism will offer him access to his mental states and the structure of his reasons; at the metalevel he will need nothing more than that access offers since the remaining principles are knowable *a priori*. For convenience, we will denote this comprehensive internalist approach as *armchair internalism*.

Advantages of armchair internalism

Armchair internalists enjoy two metaepistemic advantages over those who fail to observe the constraint on metalevel information. The first, more convenient than critical, is that they can make a clean and principled division between the object level and the metalevel, a division that externalists and even less thoroughgoing internalists are unable to maintain. This issue arises in BonJour's treatment of perceptual knowledge discussed in the previous chapter. When BonJour requires oak tree experiences to provide an "objectively good reason" for believing that an oak tree is present, his question is whether non-veridical oak tree experiences are rare. But if this sort of information is relevant to determining whether S's evidence for oak tree beliefs is "objectively good," then it will belong at the metalevel.[15] There is something quite odd about locating propositions about the presence of oak trees in various times and places at the metalevel (in order to support the claim that S has an "objectively good reason" now for believing that an oak tree is present), although these facts have no place at the object level for this particular belief. In some other context, where S is reasoning from the presence of oak trees to some further proposition (e.g. the statement that he will find an acorn), the very same propositions will belong at the object level. By insisting that there be no new empirical information at the metalevel, the armchair internalist can avoid situations in which perfectly ordinary factual claims that are not about a subject's own epistemic states appear unexpectedly at the metalevel simply because they happen to be relevant to the belief-forming process he is following in a given case. In

excluding new empirical information from the metalevel, the armchair internalist isolates what is distinctive to the metalevel: this is the place where fundamental questions of rationality and epistemic status are settled.

Second and more significantly, armchair internalists are impervious to a dilemma that externalists can otherwise try to press. Roughly, the dilemma requires the internalist to choose between (1) requiring extensive metalevel knowledge, including (for inferred beliefs) knowledge of probability theory, deductive inference rules, and the like, on the part of all subjects in order for them to have object-level justification and (2) dropping any objection to the externalist practice of regarding a subject as justified in virtue of the truth of some empirical proposition (such as a proposition about reliability) of which the subject is entirely unaware.[16]

The first alternative is distinctly unappealing. It would, counter-intuitively, rule out justification for all untrained or unreflective subjects, including many epistemologists. It also seems to violate the vague but compelling intuition that there is something epistemically important about the distinction between the metalevel and the object level. If the subject must have extensive metalevel knowledge in order to have object-level knowledge, the distinction between the two levels becomes merely a matter of conceptual clarity without any significant epistemic implications. Finally, a requirement for metalevel knowledge can in some formulations generate an incoherence.[17]

But the second alternative is obviously unacceptable to internalists. It is a key feature of internalism that the subject does at least need to be justified in believing all *external-world, empirical* propositions that are relevant to his justification. If the reliability of S's watch is relevant to his being justified in believing that it is three o'clock, then it seems to a philosopher with internalist sympathies that S must have some reason for trusting his watch. Internalists balk when externalists claim that the external reliability of some belief-forming practice can justify the subject even if the subject knows nothing about it. If the reliability of some source is required for a subject to be justified, the internalist wants the subject to have reason to trust that source. Richard Fumerton puts the point well as regards astrology: if Mother Paula the astrologer informs us that she can tell from the positions of the stars that a year of prosperity lies ahead, we will be inclined to say that she is unjustified *simpliciter* (not simply unjustified in believing herself justified) if she has no reason whatsoever for thinking that stellar configurations are correlated with economic success.[18]

Michael Bergmann has attempted to frame this objection to externalism as what he calls the "Subject's Perspective Objection":

> The externalist proposes an analysis according to which S1's belief that *p* is justified so long as there exists something X contributing to that belief's justification (where X is something like a reliable belief-forming process leading to the formation of the belief in question). The idea is

> that S1's belief is justified in virtue of X's existence even if S1 isn't aware of X – even if S1 doesn't conceive of X as something relevant to the appropriateness of her belief that *p*. Now I'll grant [says the proponent of this objection] that if someone else, say S2, were aware of X and conceived of it *as* a contributor to the justification of S1's belief that *p*, then *S2* would have a reason for thinking that *p* is true. But this doesn't at all suggest that *S1's* belief that *p* is justified. True, X is relevant in some way to the appropriateness of S1's belief. But this could be so even if S1 didn't conceive of X as being in any way relevant to the appropriateness of her belief that *p*. In such a case, it would – from S1's subjective perspective – be an accident that her belief is true.[19]

In response, Bergmann urges that an internalist who attempts to press this objection confronts a dilemma: either he will have to require full metalevel knowledge for object-level knowledge as part of his own position or he will have to submit to a *tu quoque* – to admit that he, like the externalist, allows a subject to be justified "in virtue of" some factor of which the subject is not aware.

But an armchair internalist can object to externalism without requiring that the subject have knowledge of probability theory and the like for object-level knowledge. Employing the distinction we made in the last chapter between inference forms and general "belief forming practices," he can grant that the subject reasons in accordance with inference forms and that doing so correctly from justified foundational beliefs renders his inferences justificatory. There is at least *prima facie* plausibility to the position that the subject does not need to believe the *a priori* axioms of probability or the rules of deduction in order to be justified by following an inference that correctly exemplifies these inference forms.[20] On the other hand, if a subject draws conclusions about the future from tea leaves, it is perfectly legitimate for us to demand that he have some reason for thinking that tea leaves are predictive of the future. There is no such thing as "reasoning in accordance with a tea leaf inference form." Any connection between tea leaves and the future is a contingent, empirical matter; therefore, the subject's object-level argument is not rationally cogent unless it includes a justified belief that such a connection exists.

The internalist does have an objection to externalism, but Bergmann's exposition fails to capture it. Bergmann attributes to the internalist the notion that the subject must know about anything that "exists" and "contributes to a belief's justification." But this wording is ambiguous precisely as between the "contribution" made by an object-level premise and the "contribution" of a metalevel fact about, for example, the intrinsic connection to truth that the subject's belief enjoys.

Once we observe level distinctions and confine reliability claims to the object level, the internalist's objection can be seen much more clearly. The

mere fact that some belief is reliably formed will not, according to the internalist, mean that the belief is justified. In some cases, if the subject were justified in accepting the reliability of some source and did believe it to be reliable, this belief could contribute to some other belief's justification; but it would do so *as the premise of an object-level argument*.

This clarifies the internalist analysis of the astrologer. From looking at the stars, Mother Paula is justified in believing that prosperity lies ahead only if she infers that conclusion by an object-level argument from information that leads her reasonably to trust astrology. Her object-level justification thus depends upon knowledge about the reliability of astrology, but not because metalevel knowledge is required for object-level knowledge.

With these distinctions in place, the armchair internalist can object that the externalist erroneously places statements about extrinsic truth connections at the metalevel and treats their truth as sufficient for justification. The internalist challenges these aspects of externalism not because he has a general requirement that the subject know about any "thing" that "exists" that in any sense "contributes" to his belief's justification but rather because he requires that external-world empirical information relevant to justification be present and itself justified at the object level. This requirement is itself one of the key points at issue in the dispute between internalists and externalists. But by clarifying the point at which the internalist objects to externalism we do more than merely highlight a clash of intuitions; we show that the armchair internalist, who places no new empirical information at the metalevel, is not open to Bergmann's attempted *tu quoque*.

Epistemic circularity and the metaregress

Few philosophers have worked as hard as William Alston to quiet concerns about epistemically circular arguments. In his original essay, Alston never gives a brief, clear definition of epistemic circularity, although he distinguishes it explicitly from logical circularity, in which "the conclusion actually figures among the premises."[21] In contrast, he says that in making an epistemically circular argument, one shows that one is "practically" assuming or presupposing the conclusion of the argument to be true.

> What all this comes down to is that in using or taking this argument to establish [its conclusion], one is already, implicitly or explicitly, taking [that conclusion] to be true. In this way we might say that the argument "presupposes" the truth of the conclusion, although the conclusion does not itself appear among the premises. Note that the necessity of this presupposition does not stem from the logical form of the argument, or from the meaning of the premises. . . . It stems rather from our epistemic situation as human beings. . . . Thus the presupposition falls into the large basket called "pragmatic". More

> specifically, we might call it an "epistemic" presupposition, since it depends on our epistemic situation *vis-à-vis* [the subject of the argument]. In parallel fashion we might term the kind of circularity involved "epistemic circularity."[22]

More recently, Alston gives a clearer and more succinct explication of this form of circularity.

> Since this kind of circularity involves a commitment to the conclusion as a presupposition of our supposing ourselves to be *justified* in holding the premises [for that conclusion], we can properly term it 'epistemic circularity.'[23]

Alvin Plantinga gives a description of epistemic circularity that casts some light on the concept, although it is not an entirely general definition.

> The problem with arguments for the reliability of SP is typically what [Alston] calls epistemic circularity, a malady from which an argument for the reliability of a faculty or source of belief suffers when one of its premises is such that my acceptance of that premise originates in the operation of the very faculty or source of belief in question. If you give an epistemically circular argument for the reliability of a faculty, then you rely on that very faculty for the truth of one of your premises. An obvious example would be arguing that your intuitive arithmetical faculties are reliable by pointing out that your arithmetic intuitions seem to you to be intuitively sound.[24]

This explanation is too narrow in two ways. First, it presupposes a reliability requirement. An adequate definition should not assume that epistemic circularity arises *only* from a reliability requirement. Although reliabilism is closely connected to epistemic circularity, other external requirements, e.g. that a design plan be operating in its proper environment, can generate epistemic circularity as well.

Second, Plantinga's discussion applies only to inferred beliefs, for foundational beliefs will have no premises. Yet even there epistemic circularity may arise. Within Plantinga's own system, for example, the proposition "God exists" may be held as "properly basic" without any premises. If one were to defend the claim that one is justified (or, in Plantingian terms, "warranted") in holding it, using Plantinga's own theory, one would state, *inter alia*, that God has designed us to have non-inferred spontaneous beliefs in His existence. As in Plantinga's example of mathematical intuition, the proposition to be believed appears at its own metalevel. And although Alston, like Plantinga, speaks of taking oneself to be justified (or warranted) in holding *premises* when he describes epistemic circularity, the example of Plantingian belief in God also has affinities with epistemic circularity as

Alston explains it, since it involves taking a proposition as a premise for the claim that one is justified in holding that proposition.

We shall therefore say that *epistemic circularity is present whenever p appears in the defense of an ascription of epistemic status to p.*[25] There are, then, two types of epistemic circularity, one applying to arguments and one to foundational beliefs. First,

> an argument for *p* is epistemically circular in relation to a metalevel defense of *Jp* iff *p* appears at the metalevel as part of the defense of *Jp.*

Second,

> the holding of a foundational belief *p* is epistemically circular in relation to a metalevel defense of *Jp* iff *p* appears at the metalevel as part of the defense of *Jp.*

Alston's and Plantinga's examples are clearly of the first type. If it is necessary to use *p* in order to defend the claim that one of the essential premises of the object level argument for *p* is justified, it is necessary to use *p* in order to argue that *p* is justified.

It is worth noting that no object level argument is *intrinsically* epistemically circular. This is only to be expected, given that epistemic circularity is not logical circularity. The circularity involved arises in the interaction between metalevel and object level; hence any definition of epistemic circularity, and any explanation of how it arises in a specific case, requires reference to the metalevel.

One may, as a manner of speaking, refer to an argument as epistemically circular without making any explicit reference to the metalevel. But this makes sense only as a concise way of saying that such an argument *must be* epistemically circular in relation to its own metalevel – that is, that an epistemically knowledgeable person defending the claim that *p* was justified or warranted would *have to* use *p* as part of that defense. Interestingly, whether one takes an argument to be in this sense necessarily epistemically circular will depend upon one's epistemic theory. For example, if one has a reliability requirement for justification, then one must hold that an inductive argument for the reliability of induction is of necessity epistemically circular, because the argument itself confers positive epistemic status only if its conclusion is true. Hence the statement that the conclusion is justified can be defended only by way of the claim that the conclusion is true. But if one does not insist on a reliability requirement for justification, one could argue inductively that induction is reliable without thereby using an epistemically circular argument. The relevant metalevel claim could instead be that it is *rational* to follow induction, not that induction is reliable, so interlevel circularity need not arise. Similarly, the Plantingian properly basic belief in God is, given Plantinga's own theory, epistemically circular because

an epistemically knowledgeable person would need to defend the epistemic status of the belief in part by reference to the existence of God, but this might not be the case for that same belief on a different epistemic theory.

We can now define the *metaregress* as a hierarchy of defending arguments for isomorphic claims at each ascending metalevel – e.g. that *Jp*, *JJp*, *JJJp*, etc. The metaregress may be infinite in one of two ways. Either the defense at each level involves new premises not introduced on earlier levels and the metaregress continues infinitely with ever-new arguments, or epistemic circularity arises between two of the levels. In the latter case, the metaregress can be considered infinite because the loop of epistemic circularity is repeatable *ad infinitum* and hence the conditional "*p* is justified if . . . " is never fully discharged. Briefly put, if epistemic circularity has arisen, there will be an infinite regress involving isomorphic claims (e.g. about the reliability of some belief source) at each new level. Alston regards epistemic circularity as inevitable and the possibility of an infinite metaregress involving new premises at each level as dismissible on the grounds that, given human finiteness, there will be only a finite number of "belief sources" to which justification arguments can refer.[26] For our purposes, there is no epistemically important difference between a regress involving an infinite number of type-distinct claims at each level and an infinite repetition brought about by epistemic circularity.[27]

From externalism to the metaregress

An infinite metaregress arises immediately from a requirement for a contingent R-property statement. And this situation arises whenever the externalist does not grant that, necessarily, a given belief has positive epistemic status solely in virtue of the satisfaction of internalist requirements. Since the connection to truth in such cases is contingent, all propositions constructed on the model of X (the R-property statement) are contingent claims. Hence, the reliability requirement applies to them as well.[28] A subject's belief that X itself has positive epistemic status only if some contingent claim similar to X is true of X. Therefore some contingent statement like X will have to appear as part of a defense of the positive epistemic status of X, and so *ad infinitum*. Hence any argument placed within a distinctively externalist epistemic context will either be epistemically circular or generate an infinite metaregress. And the defense of the distinctively external epistemic status of any foundational belief will also require a contingent statement to the effect that its source has the relevant R property or properties, again generating a metaregress. And as we noted earlier, on some externalist models even *a priori* beliefs, including *a priori* beliefs about epistemic status, can generate such a metaregress, since these beliefs themselves do not *of necessity* satisfy the externalist requirements. Therefore, any claim about their positive epistemic status requires a defense that includes an R-property statement, and once again a metaregress ensues.

Externalists who hold this kind of position about the *a priori* will immediately apply a *tu quoque* to armchair internalism, arguing that the metaregress cannot be stopped in any event.[29] In Chapters 5 and 6 we shall respond to this claim by articulating and defending a view of the *a priori* that stops the metaregress by precluding the need for a defense of the epistemic status of truths known *a priori*. Here we will only point out that the requirement for a contingent R-property statement *guarantees* the continuation of the metaregress, whereas armchair internalism at least holds out the hope of stopping it.

But first we must inquire whether the infinite metaregress is epistemically vitiating or whether, despite appearances, the externalist can concede all that has been said without admitting anything damaging to his position. To this issue we turn in Chapter 4.

4 What's Wrong with Epistemic Circularity

The threat of skepticism shapes externalist epistemology in a distinctive way. Internalists typically want to construct arguments that satisfy many of the traditional skeptical demands, e.g. the demand for an argument for belief in an external world that starts from internally accessible foundations. Externalists, on the other hand, are typically determined not to accept those demands in the first place. They therefore try to define 'knowledge' and/or 'justification' in a way that allows the epistemologist to help himself to double handfuls of robustly commonsensical propositions regarding our immediate surroundings, and sometimes much more. In this context, the worst thing one can say to an externalist is that his position leads to skepticism, since avoiding skepticism is one of the driving forces behind his entire project.[1]

We do intend to say that externalism entails skepticism, yet we do not mean by this what an externalist would mean were he to level the same charge against an internalist. An externalist might argue that the internalist, accepting impossibly difficult standards for positive epistemic evaluation, must say that we *in fact* do not know many ordinary things that we habitually take ourselves to know. But ironically, the externalist's flight from skepticism takes him to a more drastic position that we will term 'epistemic nihilism,' the position that *necessarily* there is no such thing as knowledge. This is a strong indictment, and making the case for it requires some care.

Internalism, externalism, and higher level requirements

In the previous chapter we have argued that externalism, as we define it, entails the existence of an infinite metaregress for those beliefs that have their epistemic status in virtue of satisfying distinctively externalist norms. As we also mentioned, William Alston has already noted this fact, yet he advocates externalism and hence clearly does not think this a damaging criticism. Alston's willingness to accept a metaregress without agreeing that it destroys the possibility of positive epistemic status arises from his conviction that this consequence would follow only if one accepted what he

calls "higher level requirements" for justification. A rejection of higher-order constraints on knowledge, Alston claims, renders the metaregress harmless, in the sense that without such constraints epistemically circular arguments or arguments that set off an infinite metaregress can nonetheless be justificatory.

> [S]urprisingly enough . . . [epistemic circularity] does not prevent our using such arguments to *show* that sense perception is reliable. . . . Nor, *pari passu*, does it prevent us from being justified in believing sense perception to be reliable by virtue of basing that belief on the premises of a simple track record argument. At least this will be the case if there are no "higher level" requirements for being justified . . . such as being justified in supposing the practice that yields the belief to be a reliable one, or being justified in supposing the ground on which the belief is based to be an adequate one.[2]

The most obvious examples of such higher-level requirements are a KK thesis – that S knows that *p* only if he knows that he knows that *p* – and a JJ thesis – that S is justified in believing that *p* only if he is justified in believing that he is justified in believing that *p*. A third, weaker, principle (which resembles Alston's last-cited example) is Richard Fumerton's "principle of inferential justification," according to which S is inferentially justified in believing that *p* only if he is justified in believing that his premises make *p* probable.[3]

Both the KK thesis and the JJ thesis seem to require the holding of an infinite number of beliefs as a *condition* for justification or knowledge – a requirement that would make justification impossible. Fumerton has pointed out that on one interpretation these principles involve incoherence. The philosopher who holds such a thesis supposedly has some finite set of sufficient conditions for justification, but the addition of a JJ requirement as another *condition* means that his initial set of conditions was *not* actually sufficient. One can hold these theses and avoid these problems if one makes the useful distinction (pointed out by Fumerton) between the idea that a second level belief (e.g. the belief that one is justified in believing that *p*) is partly *constitutive* of the justification for *p* and the idea that being justified in believing that *p* does in fact *entail* that one is justified in believing that one is justified in believing that *p*. In the latter case, one's initial (finite) set of conditions can indeed be sufficient for justification, and it may be possible to satisfy all of them. But even when a JJ or KK thesis is construed in this more careful fashion, it seems to some philosophers that either one entails that everyone who has a justified belief has an infinite number of *increasingly complex* other beliefs.[4]

A more straightforward objection is that it simply does not seem to be *true* in all cases that being justified in believing that *p* entails being justified in believing that one is justified in believing that *p*. To be justified in believing that one is justified in believing that *p*, one must at a minimum

have the concept of epistemic justification; but possessing this *concept* is not a necessary consequence of being justified in believing a proposition, unless the proposition is itself about justification. This point goes beyond the psychological fact that most epistemologists *want* to be able to say that unreflective subjects can be justified; analytically speaking, there simply is nothing about justification *per se* that means that having the concepts required for a higher-level belief is a concomitant of being justified in an object-level belief.

The "principle of inferential justification" is a more difficult claim to dismiss. Even if it is taken to state a constituent of inferential justification, it need not generate an infinite metaregress of requirements for justification so long as it is possible to have, at the metalevel, non-inferential justification for the principles of reasoning one is following at the object level. Nonetheless, it seems to us either too strong to be plausible or too weak to be valuable. A great deal depends upon how strictly one construes the notion of "being justified in believing that one's premises make one's conclusion probable." On a very strong construal, the principle would require subjects who draw conclusions using, say, Bayesian reasoning to know Bayes's Theorem, and those who reason inductively to possess in its entirety a convincing defense of the rationality of induction (such as we try to provide in Chapter 7), a requirement that is unmotivated, to say the least.

On a loose construal of the principle of inferential justification, more or less inexplicit metalevel understanding is a direct consequence of – and perhaps part of the analysis of – the act of inferring or basing itself.[5] But if one takes the principle in this way, it is difficult to see why one should make such understanding a requirement for justification. A subject's fuzzy notion that his premises make his conclusion probable is not merely fallible, but (one would think) extremely shaky; in most cases it would not be sufficiently clear even to make him rational in holding the metalevel belief. And there is no principled way to decide just *how* clear a subject's metalevel ideas must be in order to satisfy this requirement, no obvious stopping point between absolutely clear and distinct knowledge of all transmissive principles and a vague feeling that one's premises are somehow helpful to one's conclusion.

In any event, there is a distinction between reasoning *in accordance with* particular inference forms and using the metalevel claim that the inference form is valid or preserves rational confidence as a *premise* of the object-level argument. Indeed, it would be a mistake to bring such metalevel claims in as premises at the object level. Hence there seems no reason to deny that one can reason in accordance with rules of inference without making *any* use of actual beliefs about those rules. It is worth noting that the alternative to a subject's believing that his premises make his conclusion probable need not be his *denying* this proposition (and hence, by his own lights, being *irrational* in drawing his conclusion) but may be his lacking the concepts necessary to have or to evaluate the metalevel claim.

The most plausible arguments for the principle of inferential justification concern inferences like the astrological example used in the previous chapter. An astrologer may claim that a year of prosperity lies ahead because of the positions of the stars. Fumerton insists, quite rightly, that one can rationally make this inference only if one is justified in believing a proposition making a relevant connection between the positions of the stars and prosperity – e.g. "When the stars are in such-and-such positions, a year of prosperity usually lies ahead."[6] (At least, this will be true given most plausible sets of background information.) Similarly, if one concludes from the color of litmus paper that a solution is acidic, one is generally not justified in this inference unless one has reason to believe that red litmus paper usually indicates an acidic solution. But appearances to the contrary notwithstanding, these examples actually do not require one to know that the premises "make the conclusion probable" in the metalevel sense of "probable." What one needs to know in these cases is an empirical claim about class ratios – e.g. the ratio of cases in which litmus paper is red to the cases in which the solution is acidic. This is not a claim about epistemic probability, i.e. rational degrees of credence; it is rather an ordinary empirical object-level premise in one's argument. Hence it need not be taken to be a metalevel claim at all. The argument form in these cases can be induction, a direct inference of the sort we discuss in Chapter 7. Nothing in these examples tends to support a general requirement that the subject know *a priori* principles about inference forms, and the resolute armchair internalist will not be tempted to think of these cases as requiring metalevel knowledge.

We therefore do not accept a JJ thesis, a KK thesis, or the principle of inferential justification. However, this does not mean that we are unable to press the metaregress as an objection to externalism. For it is possible to reject epistemic circularity on the basis of a weaker level-connecting principle that we call the Modal Principle:

> MP: If it is in principle impossible to show decisively that S's belief that *p* is justified, then S is not justified in believing that *p*.

"Show decisively" is intended to have technical import. It refers to showing in a way that stops the metaregress that a belief is justified. Hence,

> one has *shown decisively that p is justified* iff one has exhibited a hierarchy of meta-justificatory trees such that every hierarchy terminates, in a finite number of levels, with a tree including only claims about the justification of which there can be no rational doubt.[7]

We shall refer to claims which can be shown decisively to be justified as "vindicable."

The modal principle is strictly weaker than a JJ thesis, and in two ways. First, the modal principle does not state that the knowing subject must *have*

positive metaknowledge, but only that decisive metaknowledge regarding the subject's claim must be *possible*. For this same reason, it is weaker than the principle of inferential justification. Second, the modal principle does not require that the *same subject* who holds the object-level belief must be able, given his actual concepts and abilities at t, to vindicate that belief decisively. If we adopt the modal principle rather than the other proposed level-connecting theses, we can retain the intuition that knowledge is possible for subjects who lack the intellectual capacity or the training necessary to defend a set of criteria for justification. As long as it is possible to vindicate a subject's belief, the belief passes the test of the modal principle.

The modal principle as stated concerns only justification. An externalist who favors some other term of positive epistemic appraisal instead of "justification" may insist that the modal principle is irrelevant to his position, since he is concerned with, say, warrant instead of justification. And in fact, since Plantingian warrant can in many cases consist entirely of external conditions, all of these conditions can *in fact* be satisfied even if it is not possible for anyone to vindicate the claim that they are satisfied. We shall therefore argue for a stronger modal principle (the Strong Modal Principle or SMP), of which MP is intended to be a special case.

> SMP: For any term E intended to indicate positive epistemic status, if it can be the case for some belief *p* that *Ep* while it is not in principle possible to show decisively that *Ep*, then E is not in fact a type of positive epistemic status.

For SMP, "show decisively" will also have its import with relation to E. Thus:

> For any term E intended to indicate positive epistemic status, one has *shown decisively that Ep* iff one has exhibited a hierarchy of metalevel trees such that every hierarchy terminates, in a finite number of levels, with a tree including only claims for which it is not possible rationally to doubt that they have the property E.

MP is a special case of SMP because justification is by definition a form of positive epistemic status. In Chapters 6 and 7 we will give examples of principles of reasoning that meet the condition of rational indubitability.

Could an externalist reconcile his position with this requirement? It might seem that, if a subject's belief fulfills a reliability requirement, a sufficiently knowledgeable person (an omniscient being, for example) would know this fact and hence would be able to satisfy the modal requirement for that belief. And this would seem to be the case even if it is only a contingent, empirical fact that the belief has the relevant R property, since the omniscient being would know empirical facts as well as *a priori* truths. On this view, the metaregress is just a consequence of our limited human

knowledge about empirical states of affairs. If this were the case, the metaregress would be an accidental rather than a necessary result of externalism and hence might seem less damaging. Furthermore, we have said that the metalevel can be considered from the perspective of an "ideal observer" who fully grasps epistemic standards and also has complete access to the subject's mind. Perhaps someone would argue that we should in fairness consider the externalist metalevel from the perspective of an *externally* "ideal" observer who knows any empirical facts relevant to S's warrant or externalist justification.

In fact, however, even an omniscient being would not be able to end the metaregress given externalism. Let us suppose that a subject holds a particular empirical belief and that an omniscient being knows that the subject's belief possesses the required externalist connection to truth. Nonetheless, the omniscient being's belief about the relevant characteristics of S's belief will itself either satisfy or fail to satisfy any given externalist reliability requirement, and the R-property statement about that belief will have to be made at the metalevel relative to that belief. That the omniscient being's own belief has positive epistemic status is open to rational doubt, even given that the omniscient being is aware of the relevant empirical facts about S's belief-forming mechanisms, for the omniscient being's belief about S's belief has its own positive epistemic status only if some empirical condition holds. Hence the R-property statement pertinent to the omniscient being's belief will have to appear as part of the defense of the claim that the omniscient being's belief has positive epistemic status. The omniscient being may also know *that* contingent fact, but even if he does, the positive status of that belief will require defense by means of a contingent statement at yet another level, and so *ad infinitum*. Hence, even a being who knows all empirical facts cannot stop the metaregress engendered by externalism.

Both Alston and Plantinga address this question and admit that, on their theory, an omniscient being (who has beliefs) cannot decisively vindicate those beliefs. Alston's one qualification is that God does not have beliefs and that therefore "this particular problem does not come up for God."[8] But that reservation does nothing to mitigate the unsavory fact that, under externalism, *no one* can stop the metaregress. Plantinga seems almost pleased about this particular consequence, seeing in it support for his cheerful Reidian attitude:

> Not even God himself, necessarily omniscient as he is, can give a noncircular argument for the reliability of his ways of forming beliefs. God himself is trapped inside the circle of his own ideas. About all we can say about God's ways of forming beliefs is that it is necessary . . . that a proposition *p* is true if and only if God believes *p*. Of course God knows that and knows, therefore, that all of his beliefs are true. However (naturally enough), he knows this only by virtue of relying on his ways of forming beliefs. If, *per impossible*, he became a bit

> apprehensive about the reliability of those ways of forming beliefs, he would be in the same epistemic boat as we are about that question. He couldn't give an epistemically noncircular argument for the reliability of his ways of forming beliefs; for the beliefs constituting the premises of any such argument would themselves have been formed in those ways. But any epistemic debility that afflicts a necessarily omniscient being is hardly worth worrying about.[9]

The argument here can with more justice be inverted: a theory according to which an omniscient being cannot satisfactorily answer (even his own) skeptical questions about the epistemic status of his beliefs is highly suspect on its face.

Externalism as we have defined it is therefore incompatible with the adoption of the strong modal principle. Nevertheless, it might seem that something important to internalists will be lost if internalists adopt the strong modal principle in lieu of the other proposed level-connecting principles. If the subject himself does not have to know any metalevel facts in order to have internalist justification, what becomes of the internalist concern for the subject's access to information relevant to his justification?

This question *may* arise from a confusion between the possession of object-level evidence and the possession and full comprehension of epistemic concepts. Internalism requires that the justification of a non-foundational belief is a function of the evidence to which the subject has access. And for a foundational belief, the characteristics of the belief in virtue of which it is non-inferentially justified (e.g. its specific phenomenal qualities) must also be accessible to the subject. But it does not follow that the subject *as he actually is* must be able to explain how and why these factors and evidence are justificatory.

The situations in which the internalist objects most strenuously to the externalist's notion that a subject need have no access to a particular set of factors are those in which the factors in question are empirical facts, perhaps about the reliability of some source of information or about the "normal" nature of the cognitive situation. The internalist position should be that, when they are relevant to justification at all, these factors need to be directly or indirectly accessible to the subject because, in these particular cases, beliefs about them need to be part of the subject's object-level argument.

Even so, internalists may find the separation between the untrained subject and the "expert" epistemologist disquieting. It would be an epistemic disaster if the non-expert subject could be justified despite the fact that he *cannot*, in some very strong sense of that term, find out whether he is justified – for then he would be no better off than an arational being who simply "functions" correctly or incorrectly according to canons of belief-formation that are entirely inaccessible to him.

At this point a combination of *apriorism* and armchair internalism comes in handy. In an important sense, *a priori* truths should not be regarded as

"inaccessible" to a subject simply because he lacks the concepts and abilities necessary for understanding them fully. Certainly they are not "inaccessible" in the sense in which that term is usually applied to empirical facts for which the subject possesses no evidence. If there is *no new empirical information* at the metalevel and if true epistemic principles are knowable *a priori*, an "upgraded version" (so to speak) of any believing subject will be capable of evaluating decisively the epistemic status of his own beliefs. The subject needs the power of attention and the acuity to see his own object-level beliefs at t clearly, including the way in which they are held (e.g. foundationally, by direct acquaintance, inferentially, with a particular degree of credence), and to see any connections among them, including the inference forms he is following. He must have a sufficient epistemological conceptual apparatus to evaluate correctly the epistemic status of *p* at t using conceptual analysis. *But he does not need to know any additional empirical information*. There is thus not only no threat of a metaregress but also no need to give the subject contingent information which he does not otherwise possess at t.

The distinction between these two types of accessibility makes it comprehensible, if a bit unusual, to speak of an alert and cognitively sophisticated version even of a young child – "Johnny Wideawake," to borrow Watkins's felicitous phrase – as the *same subject* evaluating the epistemic status of his own beliefs at a given point in time.[10]

What's wrong with epistemic circularity

Those with some sympathy for the externalist perspective may be inclined to dismiss epistemic circularity as negligible and the SMP as a Cartesian strait-jacket. It might be tempting to conclude with Alston that the important thing is that we *be* justified. And after all, one might think, if the epistemic principles underwriting our beliefs are true ones, then we *are* justified even under the looming shadow of an infinite metaregress. We need not attempt to traverse the weary infinities of that regress if we do not wish to; and so long as we remain satisfied with just being justified, we will not wish to. What more could anyone want?

But it is not only internalists who have expressed dissatisfaction with the "justification" externalism offers. Alston himself has had some second thoughts.

> What I pointed out in the previous paragraph is that *if sense perception is reliable*, a track record argument will suffice to show that it is. Epistemic circularity does not in and of itself disqualify the argument. But even granting that point, the argument will not do its job unless we *are* justified in accepting its premises; and that is the case only if sense perception is in fact reliable. This is to offer a stone instead of bread. We can say the same of any belief-forming practice whatever,

> no matter how disreputable. . . . [W]hen we ask whether one or another source of belief is reliable, we are interested in *discriminating* those that can reasonably be trusted from those that cannot. Hence merely showing that *if* a given source is reliable it can be shown by its record to be reliable does nothing to indicate that the source belongs with the sheep rather than with the goats.[11]

Here Alston puts his finger on the difficulty. As we argued in the last chapter, distinctively externalist positive epistemic status cannot be decisively shown to obtain. The metaregress affixes an undischarged conditional to all of the externalist's arguments and even to his foundational beliefs. His beliefs have positive status *if* something else is the case, and he can defend the claim that this "something" is the case if he possesses information relevant to it. But that argument itself gives positive status to its conclusion only *if* something else is the case, and so on. Just as a circular argument at the object level precludes ending the regress of reasons by discharging the conditional "*p* if (something else) . . . ," so an epistemically circular argument precludes the possibility of ending the metaregress by discharging the conditional, "*p* is justified (warranted, etc.) if (something else) is true."

Alston nevertheless insists that epistemically circular arguments can be justificatory. Although he is an object-level foundationalist, he assumes that only the object level requires a decisive regress-stopper. His confidence on this point arises from his argument regarding the relation between object-level beliefs and metalevel epistemic principles such as this:

> EP: If one believes that *p* on the basis of its sensorily appearing to one that *p*, and one has no overriding reasons to the contrary, one is justified in believing that *p*.[12]

Alston's point regarding justification is quite straightforward: if an epistemic principle is true, and if a belief satisfies its antecedent, then the belief is justified, period.

> [W]hat my *being* justified in believing that *p* depends on is the existence of a valid epistemic principle that applies to my belief that *p*. So long as there *is* such a principle, that belief is justified whether I know anything about the principle or not and whether or not I am justified in supposing that there is such a principle.[13]

On this view, it can make no difference whether anyone has or could have a decisive way of stopping the metaregress for such a belief. And in the case of a belief justified in virtue of satisfying EP, Alston freely acknowledges that no one can do so, since EP depends for its truth upon the reliability of sense perception – a contingent, empirical fact.[14]

A distinctively externalist problem of the metaregress arises, as we have already shown, from the need to defend the claim that an extrinsic connection to truth holds when defending the epistemic status of some object-level belief *p*. But neither the R-property statement relevant to *p* nor the epistemic principle (such as EP) for which the truth of that statement is a necessary condition makes any reference to the *particular* object-level belief *p*. Even if one is not, for example, arguing that *p* was *in fact* produced by sensory perception, externalism engenders a metaregress. Externalism, therefore, violates not only SMP but also a related principle:

> SMP′: For any term E intended to indicate positive epistemic status, for any putative epistemic principle PR stating sufficient conditions for E, and for any belief *p*, if it is possible that *p* has E solely in virtue of its satisfying the antecedent of PR, while it is not in principle possible to show decisively that the satisfaction of the antecedent of PR is indeed sufficient for E, then PR is not in fact an epistemic principle and E does not in fact denote a type of positive epistemic status.[15]

For example, it is possible for some belief *p* to have Plantingian warrant even though it is not possible to show decisively that a particular design plan module (the module, such as sensory perception, that in fact produced *p*) is reliable when functioning properly in its intended environment. Any belief about the reliability of that module might itself be produced by a module of a design plan that was not reliable even when functioning properly in *its* intended environment. So it would not be possible to defend decisively a principle like:

> Q: If a belief is produced by the human faculty of sensory perception functioning properly in its intended environment, that belief has warrant,

since a rational question could arise regarding the warrant of Q, setting off the metaregress. Yet if sensory perception does satisfy the Plantingian requirements, a belief could have warrant solely in virtue of satisfying the antecedent of Q. Therefore, by SMP′, Q is not an epistemic principle and 'warrant' does not denote a type of positive epistemic status.

An externalist might try to circumvent this problem by adopting an epistemic principle of much greater generality than Q, arguing that no R-property statement is necessary for the support of that broader principle. For example, he might state as his epistemic principle:

> Q′: Justification supervenes on reliable belief production within the proper design plan

and construe this as a necessary truth knowable *a priori*. No empirical information will then be required for the support of Q′, so the externalist appears to have escaped the net.

But in that case, the R-property statement relevant to *p* (e.g. "Sensory perception is a reliable source of beliefs") would not disappear or become irrelevant; it would have to be brought in when discussing the question of whether S's belief was reliably produced. Because SMP′ places strictures on all putative epistemic principles, these two different ways of bringing the R-property statement into play make no difference to our case. For it would always be possible to construct a more specific externalist principle, such as Q, that was relevant to *p*.

Since the armchair internalist has no new empirical information at the metalevel, his epistemic principles will pass the test of SMP′ regardless of their degree of generality. For example, both the statement, "If S's premises confer upon his belief that *p* a high degree of rational credibility, then S is justified in believing *p*," and "If S's premises confer upon his belief that *p* a high degree of rational credibility by way of inductive extrapolation, then S is justified in believing *p*" are, according to the armchair internalist, necessary truths.[16]

What we must ask, then, is whether statements like EP and Q can be real epistemic principles even if they cannot be defended in a way that stops the metaregress. Although Alston takes himself to have argued that epistemically circular arguments can be justificatory, he has not argued directly for this point. His argument that epistemic circularity does not preclude justification turns on a handful of distinctions that Alston thinks internalists are prone to overlook, particularly the distinction between "being justified" and actively "justifying."[17] Elsewhere we have examined these arguments in detail and found that none of them does the work Alston needs it to do and that the controversy therefore comes back once more to the question: can a genuine epistemic principle be invindicable?[18]

One way to see the force of SMP′ is to consider the epistemic constraint that propositions about which a rational question could arise but that are in principle not defensible cannot bear epistemic weight.[19] Suppose, for example, that it is completely impossible to tell (due, let us say, to empirical equivalence and an inscrutable balance of theoretical virtues) whether there is absolute spacetime. Then the proposition, "There is absolute spacetime" cannot be defended in a way that could convince someone who was skeptical about it. Regardless of whether such a proposition is true, if it really is indefensible, it cannot be used as a premise to justify belief in anything else. If it is indefensible, it is epistemically irrelevant.

To apply a similar intuition to the metalevel we must consider the special nature of epistemic principles. By their very nature, epistemic principles purport to have epistemic implications. If an epistemic principle is genuine, this entails that object-level propositions satisfying its conditions have positive epistemic status in virtue of satisfying those conditions. Hence, to

claim that *there can be genuine epistemic principles that, although rationally dubitable, cannot be defended against someone who is skeptical about them* is to allow an indefensible proposition to have epistemic weight.

Why should the externalist care?

At this point an externalist can bite the bullet, and he might not even consider it a very problematic bullet to bite. The kind of epistemic weight being given to the epistemic principle is different from the kind of weight that an indefensible object-level proposition would have when used as a premise. Our lower-level beliefs are not based *inferentially* on any epistemic principle. The externalist might insist that only *that* epistemic role, the role of giving inferential support, is off limits for indefensible propositions.

But here a threat of arbitrariness arises, a threat that should concern even an externalist. If indefensible claims can be genuine epistemic principles, then there can be epistemic principles that are *internally* indistinguishable from such farcical "principles" as:

> FP: If S believes that *p* because *p* has been expressed to S by someone who wears mismatched socks, then S is justified in believing that *p*.

It does not matter whether *p* often happens to be true under such circumstances. Even if the truth of beliefs formed under these conditions were guaranteed by the continual miraculous intervention of an omnipotent deity, FP would not be a serious candidate for an epistemic principle.[20]

Yet, like any putative epistemic principle that cannot be defended in such a way as to stop the metaregress, FP is "vindicable" by way of an epistemically circular argument. S can assert FP and, when challenged about it, can support it by saying that it was told to him by someone wearing mismatched socks. (As a last resort, he can even put on mismatched socks and express it to himself.) In response to the obvious question, "So what?" S (quickly slipping into his epistemologist's hat and socks) can reply that the inference from "*p* was told to me by someone wearing mismatched socks" to "*p*" is justificatory in virtue of its satisfaction of the higher-level "principle," FP. If pressed, one could easily invent an arbitrarily large set of such principles and use them to "reinforce" each other.

By this sort of reasoning, *anything* can be viewed as "possibly justified." After all, *if* FP is a real metaprinciple, then those beliefs comporting with its requirements (including, in this case, FP itself) are justified. And if we consider it meaningful to speak of epistemic principles that are nonetheless invindicable except by way of such epistemic circularity, then we have absolutely no reason not to take FP, and the epistemically circular argument S gives for it, seriously. It could just be the case that FP is one of those invindicable but nonetheless genuine epistemic principles and therefore that beliefs satisfying it are in fact justified. On the assumption that real

metaprinciples can be invindicable, there is simply no way to tell. This, of course, is what Alston means when he says that epistemically circular arguments cannot help us to distinguish the sheep from the goats.

This problem arises because internal rationality is not necessary for positive epistemic status on an externalist view. It is obviously not *internally* rational to believe some proposition *merely* because it was stated by a person in mismatched socks. But this does not look like a damaging criticism if one adopts externalism. Even if one has no internally accessible evidence for a belief, and even if a belief falls into no internally defensible category (such as being known by direct acquaintance) that makes it foundational, there is no way for a principled externalist to judge the belief to be epistemically "bad." So propositions believed in accordance with FP have no externalist mark against them. Conversely, since internal rationality is not sufficient for externalist positive epistemic status, beliefs held in accordance with principles that are not obviously farcical – e.g. a principle that states that induction can confer justification – have no *prima facie* positive status, given externalism.

Nor can an externalist use an appeal to common sense to avoid taking FP seriously. Given externalism, it would be arbitrary to treat FP as a non-starter solely on the basis of some spontaneous feeling that it is highly implausible. If there can be real epistemic principles for which only an epistemically circular argument can be given, how do we know that FP is not one of them? On the externalist view, any sort of intuition that some principles are better than others is itself defensible only in a non-vindicable fashion. One need only ask, "How can you tell that your feeling that some EPs are absurd is *reliable*?" to see that the externalist cannot appeal to the evident silliness of the suggested principle without, on his own account, becoming mired again in the metaregress.

The externalist position thus entails epistemic anarchy: it entails that there is, *in principle*, no way of decisively distinguishing genuine epistemic principles from absurd ones and hence no way of decisively distinguishing beliefs that have positive epistemic status from those that do not. In view of this, we must regard the claim that genuine metaprinciples can be invindicable as incoherent. For on this assumption one can have an epistemic principle that is absolutely worthless, epistemically – incapable of doing any epistemic *work* in the way of showing that some beliefs, as opposed to others, really do have positive epistemic status. Yet it is obvious conceptually that the *raison d'être* of an epistemic principle is to tell us unequivocally (rather than only conditionally) that some types of beliefs do have positive epistemic status.[21] It makes no sense, conceptually, to speak of an epistemically useless *epistemic* principle. An invindicable epistemic principle cannot do the job that, by definition, an epistemic principle does.

How does this argument support SMP′? We have agreed that one can stipulate a technical meaning for a term like 'warrant' (for example) and that the conditions for warrant, as Plantinga defines it, can be satisfied even

if they cannot be shown decisively to hold. But a "principle" stating, for example, that one is warranted if one accepts the deliverances of one's senses is of no epistemic value, since we cannot tell decisively if it is true. Accepting the deliverances of one's senses might or might not actually confer warrant, depending upon whether one's sensory faculties fulfil the various Plantingian requirements. But the attempt to inquire into that empirical fact is merely part of an infinite metaregress of further similar questions. Such a principle is therefore not an epistemic principle, as it does no epistemic work to distinguish decisively the types of beliefs that have the putatively desirable property from those that do not.

Since such principles cannot be vindicated, we can attribute warrant to all sorts of things without any decisive assurance that they do have warrant. But in that case, "warrant" cannot be an important *epistemic* or *philosophical* category. The question of whether a belief possesses warrant is not the kind of thing into which we can make a distinctively philosophical inquiry, and "warrant" cannot do distinctively philosophical work. As Richard Fumerton observes, "[T]he very ease with which externalists can deal with the skeptical challenge at the next level betrays the ultimate implausibility of externalism as an attempt to explicate concepts that are of *philosophical* interest."[22]

If, as we have argued, it is incoherent to postulate genuine epistemic principles that are not vindicable, then the view that no epistemic principles are vindicable entails that there are no genuine epistemic principles. And if there are no genuine epistemic principles, then there are no justified beliefs (and therefore, if knowledge requires justification, no knowledge), since by definition a justified belief satisfies an epistemic principle or set of such principles. If there is no putative type of positive epistemic status that can be shown decisively to obtain, even when a belief does possess it, then there is no such thing as positive epistemic status.

This is why we dub the position that follows from externalism "epistemic nihilism," a deeper form of skepticism than that which the externalist is trying to avoid. The externalist attempts to sidestep skeptical worries that our ordinary beliefs are not in fact justified, but his way of doing so leads to the conclusion that justification itself is necessarily unattainable. If all epistemic principles are invindicable, all that we take to be knowledge is really an *Ersatz* state underwritten by "epistemic principles" that do nothing.

It might still be objected that there is a suspicious whiff of internalism about this entire argument. Why, it might be asked, should an externalist care if two putative principles are indistinguishable in the amount or type of *internal* evidence in favor of them? Well, does the externalist want to take FP seriously? This seems a far less palatable bullet than it may have appeared when cast in such abstract terms as "giving indefensible propositions epistemic weight." To be sure, we are, by pressing this point, trying to access whatever last vestiges of internalist intuition might remain within the externalist reader. If someone shrugs his shoulders and says, as does

Plantinga, "That's just life in philosophy," we shall have to leave him to the arbitrary selection of his epistemic principles.

Perhaps an externalist will be particularly inclined to shrug off this argument if his externalism is motivated by naturalism. He may think, with Quine, that epistemology needs to be naturalized; he may believe that metaphysical naturalism is both strongly supported by independent evidence and is incompatible with internalism.

Whether metaphysical naturalism is incompatible with internalism is a fascinating question beyond our scope in this book. So, too, are arguments for rejecting naturalism in the philosophy of mind in favor of mentalism or dualism. Our focus here is primarily on epistemology rather than metaphysics; we are presenting arguments for epistemic internalism. If epistemic internalism is indeed incompatible with metaphysical naturalism, and if there are reasons to accept internalism, this provides a reason to rethink naturalism.

But in fact, even a self-described naturalist might want to say that scientific knowledge is reflective knowledge or, at least, reflectively defensible knowledge. And it is at just that point that first philosophy becomes pertinent, perhaps initially as a kind of heckler, posing questions science cannot answer on its own behalf. As Bertrand Russell put it,

> When the behaviorist observes the doings of animals, and decides whether these show knowledge or error, he is not thinking of himself as an animal, but as an at least hypothetically inerrant recorder of what actually happens. He "knows" that animals are deceived by mirrors, and believes himself to "know" that *he* is not being similarly deceived. By omitting the fact that *he* – an organism like any other – is observing, he gives a false air of objectivity to the results of his observation. As soon as we remember the possible fallibility of the observer, we have introduced the serpent into the behaviourist's paradise. The serpent whispers doubts, and has no difficulty in quoting scientific scripture for his purpose.[23]

If scientific knowledge of the sort the naturalist reveres needs to be reflectively defensible, then the arbitrariness problem for externalism becomes acute. We need not focus on an invented principle like FP to see the arbitrariness problem in action; it is illustrated particularly well by the trouble naturalist externalists have in dismissing Alvin Plantinga. How can the naturalist externalist tell Plantinga that he cannot use "principles" concerning the reliable, God-designed faculties of human beings, including such seemingly spooky or irrational faculties as "the internal instigation of the Holy Spirit"? After all, the naturalist assumes that our faculties are reliable because they arose by way of the pressures of natural selection. This claim is itself an empirical one, intended to explain how our faculties came to be reliable and why we should trust them. But the attempt to ascend to the

level of epistemology and defend that claim as justified bogs down, given an externalist reliability requirement, in the further empirical question of whether the belief-forming mechanisms used in that investigation are properly, externally connected to truth. Richard Fumerton notes the problem for the non-theist externalist with some glee:

> [G]iven well-known externalist analyses of noninferential justification or warrant, it is surely *conceivable* that Christian belief and any other true belief is noninferentially justified or warranted. If you don't like this conclusion, don't complain to me. I'm a classical foundationalist. ... Externalists heave a sigh of relief at their ability to sidestep traditional skeptical challenges with a new, improved, expanded conception of foundations. That door's having been opened, however, Plantinga is going to barge right through and you externalists are going to have to find some non-question-begging way to distinguish your externalist dismissal of skeptical challenges to the legitimacy of memory or perception from Plantinga's attempt to exploit those same moves in defending Christian theism from the skeptic.[24]

So why should the externalist, especially the naturalist externalist, accept any version of the modal principle? Why should an externalist be concerned about the entire argument of this section or troubled by epistemic circularity? Because if he does not do something about that argument, he will be unable to give a fully principled answer to the question of why he accepts principles about, say, sense perception as justificatory but rejects those he finds repugnant, like Plantinga's invocation of the internal instigation of the Holy Spirit as justificatory. If a given externalist cannot find any problem in this scenario, he can go on his way undisturbed. But not everyone – not even every self-identified naturalist or externalist – will be able to shake the feeling that, in accepting arbitrariness at the level of epistemic principles, he has lost something of value.

The Great Pumpkin and Plantingian defeaters

The preceding argument allows us to put forward a particularly clear version of the "Great Pumpkin" objection to Plantinga's epistemology. Plantinga's entire system rests on the idea that we are "rational" when our belief-producing mechanisms are functioning properly according to a properly truth-conducive design plan. But his externalism guarantees that we cannot tell decisively and unqualifiedly that this is the case. This means that those who believe in the Great Pumpkin and think that the Great Pumpkin has designed them to believe in him cannot, however flimsy their evidence, be said decisively to be irrational or compared unfavorably with those who, even on the basis of far more evidence, believe in the Christian God. Given Plantinga's view, the epistemic status even of carefully reasoning Christians

is, so far as we can tell internally, no better than that of thoughtless or stubborn Great Pumpkinites. In this sense, Plantinga's position is, as one critic has put it, "radically relativistic."[25] Just as the naturalist has no non-arbitrary way to dismiss Plantinga, so Plantinga has no non-arbitrary way to dismiss Great Pumpkin worshippers.

Imagine a young man who comes to doubt the existence of God but eventually assembles a set of arguments – both broadly metaphysical (such as the cosmological argument) and empirical (such as the argument from the resurrection of Jesus Christ). And suppose, for the sake of the argument, that these are impeccable from the perspective of internal rationality. Is this particularly helpful to him from an externalist perspective? Can we say even now, in any decisive way, that his belief in God has externalist positive epistemic status? No. For the rational cogency of his inference forms, the infallibility of his foundational premises, even the conjunction of such conditions is not enough in itself to make his beliefs "rational" when this is understood externalistically. So if he uses inference to the best explanation or induction, and if these do not meet the extrinsic requirements, such as being reliably truth-conducive, being designed to operate in the environment in which he is using them, and the like, then his belief is still not "rational" in Plantingian terms. And if he moves up a level and investigates *whether* his belief-forming practices have those properties, he will gather more evidence and make inferences from it, and the same question will arise about the belief-forming mechanisms he uses there. So for all he can tell decisively, he might be "irrational" in an externalist sense.

On the other hand, consider Linus in the pumpkin patch, believing year after year that the Great Pumpkin will arrive, praying to the Great Pumpkin, holding fast to his faith in the Great Pumpkin even when he is repeatedly let down. Does the fact that his religious belief is obviously irrational from a traditional internalist perspective *in and of itself* mean that, on an externalist model, his belief is irrational? Of course not. Internal rationality is not necessary for externalist positive epistemic status. And it is possible that the Great Pumpkin really does exist and has designed people to believe in him despite an exceptionless record of failed predictions – to believe anti-inductively.[26] And it might be that, in the grand scheme of things, taking into account all of time and space or some relevant set of possible worlds, this means of belief formation is reliable. The fact that in Linus's case and thus far the Internal Instigation of the Great Pumpkin (if the Great Pumpkin exists) has produced a string of false beliefs isn't a devastating criticism from an externalist perspective. As Linus shouts when Lucy urges him to curse the Great Pumpkin and give up, "Just wait until next year!" So for all we can tell decisively, Linus *might* be warranted.[27]

Plantinga shows some puzzlement about the Great Pumpkin objection and tries at one point to construe it as an attack on the very possibility that voodooists or Great Pumpkin worshipers are internally rational in holding their beliefs. He points out that they might under some circumstances *be*

internally rational – if, for example, their priests or leaders were able to present the religion with great plausibility.[28] Whether this would suffice for internal rationality is questionable, but waiving that point, we may of course acknowledge that under some unusual circumstances, a person *might* have such evidence as would render him internally rational in believing in the Great Pumpkin. The point of the objection, however, is that Linus does not. Nor, we generally take it, are voodooists or astrologists internally rational in holding their views. We call UFO worship or Great Pumpkin worship or crystal ball reading "disreputable" or "crazy" because we are convinced that the people who engage in such practices are *not in fact* meeting, nor even coming close to meeting, any standards of internal rationality. The problem raised by the Great Pumpkin objection is that this fact, in itself, *does not matter* to the Plantingian, who might without blushing boast that his belief also fails utterly to meet such standards.[29]

Plantinga is well aware that internalists are unhappy with his view and that his notion of proper basicality seems to many to be a doorway for all sorts of irrational and bizarre views.[30] To criticism of proper basicality (especially in its Great Pumpkin form) he responds that, on his view, beliefs can be properly basic and nonetheless be subject to rational scrutiny, open to rational objection, and the like by way of counterevidence functioning as a defeater. Hence he is not insulating his properly basic beliefs (including his religious beliefs) from rational appraisal nor even from rejection by way of evidence.

And, a Plantingian could continue, perhaps this answers the Great Pumpkin objection straightforwardly. Does not Linus have evidence that he is *not* functioning properly according to a truth-directed design plan in the very fact that the Great Pumpkin has never appeared on Halloween night? For that matter, Linus has evidence not only that he is not functioning truth-directedly but also, and more simply, that the Great Pumpkin does not exist. And does this evidence not count as a defeater – as the kind of thing that would cause a rational individual to drop his belief? So perhaps the incorporation of defeaters into their system allows externalists some purchase in distinguishing beliefs that have externalist epistemic status from those that lack it.

In fact, Plantinga tries to use a defeater analysis to do precisely this sort of epistemic work in his evolutionary argument against naturalism. He argues that, given naturalism, the probability that one's faculties are truth-directed and reliable is either low or inscrutable. And if one's faculties are not truth-directed and reliable, then one does not have warrant for holding one's beliefs, including one's belief in naturalism. Therefore, he takes it, the naturalist has a sort of self-referential defeater for his belief in naturalism – it is irrational for the naturalist to believe naturalism itself.[31]

This approach has the odd consequence that Plantingian proper basicality automatically transfers the burden of proof from the believer to the doubter. It allows people to be externally justified (or warranted) in holding beliefs

without any semblance of internal rationality *in the first instance*, taking account of counterevidence only after the fact, as it were, by allowing that if and when counterevidence appears it can take away warrant.

But we can waive that point in favor of a more telling one. Though Plantinga uses terms such as 'rational' freely in his discussion of defeaters, discussing what a "rational" subject would do upon hearing a particular argument or learning a particular piece of evidence,[32] both the concept of rationality and the concept of a defeater are *externally defined*.[33] Plantinga's entire definition of an epistemic defeater is as follows:

> *D* is a purely epistemic defeater of *B* for *S* at *t* if and only if (1) *S*'s noetic structure *N* at *t* includes *B* and *S* comes to believe *D* at *t*, and (2) any person *S** (a) whose cognitive faculties are functioning properly in the relevant respects, (b) who is such that the bit of the design plan governing the sustaining of *B* in [his] noetic structure is successfully aimed at truth (i.e., at the maximization of true belief and minimization of false belief) and nothing more, (c) whose noetic structure is *N* and includes *B*, and (d) who comes to believe *D* but nothing else independent of or stronger than *D*, would withhold *B* (or believe it less strongly).[34]

But this definition says absolutely nothing about what an internally rational subject would do under the stated circumstances. Perhaps a subject with the stipulated type of design plan would *ignore* propositions that an internally rational subject would treat as counterevidence. If *in fact* the faculties that produced B were unreliable, or were not aimed at truth, then the subject does not have Plantingian warrant. But that is not connected at all to what the properly functioning, truth-aimed individual would do were he to come to have doubts about the reliability of his faculties. Perhaps, for example, a naturalist will actually be warranted in ignoring Plantinga's argument against naturalism. Perhaps that is what Plantinga's S* would do. The definition of an epistemic defeater does not entail that a subject has a defeater for B if he has evidence that the faculties that produced B were unreliable, much less if he merely discovers that the reliability of the faculties that produced B is "inscrutable."

And something similar can be said about Linus: On an internalist model, of course Linus would be rational to drop his belief in the Great Pumpkin after he fails to turn up. For that matter, there is a good case to be made that Linus wasn't internally rational in believing in the Great Pumpkin in the first place. But it cannot therefore be assumed that Linus has a "defeater" for his Great Pumpkin belief in the Plantingian sense. That question depends entirely upon *de facto* issues, e.g. whether there is a (in the long run) successfully truth-aimed design plan module that calls for belief in the Great Pumpkin despite a string of failed predictions of his appearance.

Thus, Plantinga's idea that counterevidence can function as a defeater *seems* to be a relevant response to internalist objections only because it trades on *internalist* concepts of rationality and counterevidence, concepts to which Plantinga has no claim. Any attempt to determine whether the properly designed, warranted individual would drop his belief in B upon learning D will set off the metaregress yet again, for it will introduce, yet again, empirical issues about what the design plan in question is, whether it is successfully aimed at truth, and the like. So, once again, externalist principles can give us no way to separate the epistemic sheep from the goats.

"Practical rationality" and "significant self-support" to the rescue?

Though Alston admits some uneasiness about epistemic circularity, he makes two attempts to mitigate the consequences of his position. First, he insists that it is "practically rational" to continue to rely on our socially well-established practices, and that it is therefore rational to believe that those practices are reliable. Alston adduces various premises in support of this claim, e.g. (1) there is no overriding consideration against using these practices, (2) there are no practices whose reliability could be established more satisfactorily than those we find ourselves using, (3) we have no choice but to use some doxastic practice or other if we are to believe anything, (4) it would be extremely disruptive and difficult to replace our current practices with others, and (5) we are socially committed to these practices.[35]

The relevance of some of these premises (particularly 5) is doubtful, even to "practical rationality," but waiving that point, this rescue of epistemically circular practices is really no rescue at all. An obvious response to Alston's suggestion comes from Plantinga himself. These premises, he gleefully points out, depend crucially for their own warrant on practices such as sensory perception and memory (and, he might have added, induction, since 4 predicts what will happen if we try to replace our practices). Therefore, granting the externalist claim that none of these practices is vindicable, Alston's argument for practical rationality is itself epistemically circular.[36]

What of Alston's claim that, since it is practically rational to use our common practices, it is rational to believe them to be reliable? This inference is very odd. How, given Alston's own position, can we conclude decisively that *any* belief is epistemically rational? But Alston eventually makes it explicit that he intends only to say that it is *practically* rational to believe our practices to be reliable.[37] This claim, while less confusing, is also far less interesting. But Alston tries to dispel any disappointment:

> If, as I have argued, we are unable to find noncircular indications of the truth of the reliability judgment, it is certainly relevant to show that it enjoys some other kind of rationality. It is, after all, not irrelevant to

> our basic aim at believing the true and abstaining from believing the false, that ... established doxastic practices constitute the most reasonable procedures to use, so far as we can judge, when trying to realize that aim.[38]

But the phrase 'most reasonable' here is itself merely pragmatic in nature, having to do with such issues as what would cause the least trouble and disruption in our lives, whether it is possible to do any better than to use what we already have, and so forth (assuming that we could know the answers to these questions). This is truly to give a stone for bread. A prisoner may have nothing to eat save what his captors set before him; this is no evidence that it will be nourishing rather than poisoning. And if the alternative is starvation, and it is therefore "practically reasonable" to eat, this reflection will not still his doubts. To be told that, in order to think at all, we have no choice but to engage in standard doxastic practices without hope of a reflective justification for them is to be told nothing of positive *epistemic* interest.

Alston's second line of argument is an attempt to make good on the deficiency of the first. For the practice of forming beliefs on the basis of sensory perception (SP), unlike crystal-ball gazing, exhibits "significant self-support."

> By engaging in SP and allied memory and inferential practices we are enabled to make predictions, many of which turn out to be correct, and thereby we are able to anticipate and, to some considerable extent, control the course of events. ... By relying on sensory perception and associated practices we are able to establish facts about the operation of sense perception that show both that it is a reliable source of belief and why it is reliable. ... Our scientific account of perceptual processes shows how it is that sense experience serves as a sensitive indicator of certain kinds of facts about the environment of the perceiver. These results are by no means trivial. It cannot be assumed that any practice whatever will yield comparable fruits. It is quite conceivable that ... this output should not have put us in a position to acquire enough understanding of the workings of perception to see why it can be relied on. ... Many doxastic practices, like crystal-ball gazing, do not show anything analogous to the above results. Since SP supports itself in ways it conceivably might not, and in ways other practices do not, its claims to reliability are thereby strengthened.[39]

This argument has a very reasonable sound to it. Indeed, it has a distinctively Bayesian ring. Without using Bayesian terminology, Alston appears to be indicating something like a disparity in likelihoods for the evidence we have as between reliable and unreliable sense perception. If sense perception were not reliable, we would not expect to find ourselves

able to predict and control our environment and able to make detailed statements about the nature of sense perception itself.

On the object level, some of Alston's points may be relevant to an argument for reliable sensory perception. Most of us do in fact believe that we are rational in accepting the deliverances of our senses, believing in an external world, and trusting scientific claims about the structure of our sensory organs. While the decisive argument against, say, Cartesian Deceiver scenarios has yet to be given in full, we are inclined to think that at least the detailed, coherent occurrent seeming memories of prediction and control will have an important role to play in the argument.

But on his own epistemic theory, Alston has cut himself off from a noncircular appeal to this sort of empirical argument. He says, for example, that we correctly predict and control events. But on his own reliabilist account, we can know these premises only if sense perception is in fact reliable. Hence, this argument is itself epistemically circular and cannot assuage doubts about epistemic circularity.

Even if he chose to cast the argument in more cautious terms, saying only that it seems to us that we have predicted accurately, etc., and even if access to our own occurrent experiences escaped the externalist metaregress, the sweeping nature of Alston's externalism guarantees that at some level this too must be an epistemically circular argument. In the final analysis, Alston would have to say that we can use Bayesian reasoning to gain justified belief only if Bayesian reasoning is actually reliable, thus starting the metaregress yet again. Empirical arguments for realism and reliable sensory access are hopeless if they are epistemically circular.

Alston does not attempt to hide the epistemic circularity of his significant self-support argument. Driven by a commendable honesty, he is forced to take away with the left hand what he gives with the right.

> Since even significant self-support exhibits epistemic circularity, I will refrain from taking it to be an independent reason for supposing the doxastic practice in question to be reliable. Because self-support requires assuming the practice in question to be a reliable source of belief, it provides evidence for reliability only on the assumption of that reliability; and that is hardly evidence in any straightforward sense. Hence I am taking significant self-support to function as a way of strengthening the prima facie claim of a doxastic practice to a kind of practical rationality . . . But as such it is by no means a negligible consideration.[40]

But if the argument is really epistemically circular, this consideration is worse than negligible; it is entirely irrelevant. How can an epistemically circular argument strengthen *any* claim, even "the *prima facie* claim of a doxastic practice to a kind of practical rationality"? Alston's answer, apparently, is that self-support is strengthening because of the contrast between

practices that seem to exhibit it and those, such as crystal-ball gazing, that do not. Here he appears to be entertaining a last hope that an epistemically circular argument can, despite his former misgivings, separate the epistemic sheep from the goats.[41] His idea is that some circles are better than others. "[T]his is not the trivial epistemically circular support that necessarily extends to every practice, the automatic confirmation of each output by itself."[42]

It is far from obvious, however, that the failure of astrology or crystal-ball consultation to produce "confirmations" of and theories about their own reliability is sufficient to establish anything favorable about sensory perception. Even if theories about a practice's reliability do provide internal evidence that the practice is in fact reliable, internal rationality is not necessary for externalist positive epistemic status. Hence, those practices could fail to provide such theories and still be entirely reliable. The absence of helpful information, coherence, detail, or any other such characteristics from crystal-ball proclamations does not in any way impugn the externalist status of the beliefs they produce, as externalism requires nothing of the sort. Similarly, the presence of evidence for the reliability of one's senses does not, *on an externalist account*, give us any decisive defense of the positive epistemic status of our sensory beliefs. For that evidence itself might lack externalist positive epistemic status, or the form of inference by which we come to the conclusion that our senses are reliable might not itself be reliable. In fact, the situation is exactly parallel to that outlined above with respect to belief in God. The presence or absence of what an internalist calls good evidence is of no help to the externalist.

Moreover, it does not matter if astrologers have not in *fact* made an epistemically circular argument for the reliability of their practice. They easily *could* do so, and it would not matter if they did. Any philosopher worth his salt would remain intransigently skeptical about astrology even if reputable astrologers faithfully reported that its reliability was affirmed or even supposedly explained *in the stars*. Nor would his concerns be allayed if there were many cases in which the astrologers predicted remote events and then told us that, according to the stars themselves, their predictions had come true.[43]

Conclusion

In Chapter 2 we agreed with Richard Fumerton that epistemologists should not allow their philosophical projects to be driven by fear of classical skepticism. It is possible that the skeptic is simply *right* about many of our beliefs. We must set up our epistemic standards without malice aforethought – without building in a guarantee that we cannot lose an epistemic right to any of our much-loved common sense beliefs.[44] Moreover, unless we have examined our concept of justification closely enough to see with absolute clarity that it is coherent, we must allow the possibility that

it is not. If we were to discover that it is incoherent, however, we should acknowledge that epistemology is at an end. We should not construct a pseudo-epistemology based on an attenuated and insignificant concept. To flout the modal principle is to resign ourselves to a counterpushing game in which we subjectively rate practices by criteria we cannot defend; in doing so, externalism eliminates *of necessity* any justification worth the name. This conclusion, ironic but ineluctable, raises the stakes considerably in the debate.[45]

5 Analytic *a priori* Knowledge

The problem of the epistemic regress described in Chapter 4 reveals a significant isomorphism between empirical and non-empirical knowledge. Empirical knowledge, as we have argued elsewhere, requires empirical foundations – justified beliefs that are knowable directly (and, on our view, with certainty) independent of any support derived by inference from other beliefs.[1] Similarly, knowledge of non-empirical truths appears to require foundations of its own if it is to escape the blight of epistemic circularity. And genuine metafoundations – metalevel propositions that stop the metaregress – will offer the Johnny Wideawake subject rationally indubitable knowledge not only of their own truth but also of his justification in believing them.

The reader who has persevered this far can scarcely have missed our repeated emphasis on the *a priori* character of metafoundations and the requirement that they be vindicable. But we have offered as yet no account of *a priori* knowledge. Here we will rectify that omission, emphasizing analytic *a priori* knowledge. By articulating a robust position regarding the epistemic grasp of conceptual truths, we hope both to show how vindicability can play the role we have assigned it and also to suggest that there is less substance to contemporary critiques of analytic *a priori* knowledge than meets the eye.

Analyticity articulated

In his sustained defense of a form of rationalism with respect to judgments, Laurence BonJour proposes two challenges for proponents of analyticity to meet if they are to sustain a claim that their analysis of *a priori* knowledge has any advantage over the rationalist's notion of synthetic *a priori* knowledge. BonJour implies that any valuable account of *a priori* knowledge must provide genuine epistemic insight as to why grasping such propositions in the indicated fashion makes one *justified*. This, he clearly believes, his account of the synthetic *a priori* is able to do, but he repeatedly questions whether analytic accounts are able to fulfill this requirement. Moreover, since advocates of analytic *a priori* knowledge have sometimes suggested

that there is something altogether too vague and mysterious about the notion of a non-analytic insight into necessary truth, BonJour challenges them to provide an account of their own that does not (to at least as great an extent as a concept of the synthetic *a priori*) leave our grasp of such truths *mysterious*.[2]

Though these requirements seem innocent enough, they are somewhat vague, and this vagueness allows BonJour considerable latitude in declaring that they have not been met by various theories of *a priori* knowledge that utilize the concept of analyticity. The first step toward demystifying analyticity is to sharpen up BonJour's challenges by giving an account of the notion of justification that can reasonably be employed in this context. This, in turn, will allow us to see whether the charge of "mysteriousness" can legitimately be leveled against the position we advocate.

An obvious way to approach the question of justification is to offer, as a sufficient condition for the justification of belief in a necessary truth, its derivation from an accepted basic set of necessary truths. This accords well with mathematical practice, but we shall not offer this analysis, for three reasons. First, in order to be satisfying, such a proposal presupposes that there is an independently certifiable set of necessary truths; and in that case, on pain of circularity, the proposal of justification by derivation cannot be the *whole* story about our grasp of necessary truth, or even the heart of the story. Second, any derivation will involve the use of principles of reasoning; and our grasp of those principles is likewise left unexplained in this account. Such an approach would therefore not be helpful in showing how *a priori* knowledge can stop the metaregress, which is our ultimate goal in this chapter. Third, the account captures only a very limited set of necessary truths. It applies mainly to the sorts of reasoning we do in mathematics and symbolic logic, but it leaves out of account a wide range of important cases where we are convinced of a necessary truth – convinced that it is true, and often at the same time and (so to speak) in the same intellectual act convinced that it is necessary – independent of any chain of reasoning. Our understanding in these cases may be, and frequently is, positive and direct. One may well first recognize that a particular proposition is necessarily true and only subsequently (if at all) analyze the conceptual connections among its terms and elicit meaning postulates. Indeed, a principal criterion of adequacy for such a system of meaning postulates is that they collectively entail the epistemically pellucid judgments with which we started.[3]

None of these objections is telling against the project of formalization as such. It would be absurd to suggest that the advances in logic over the past century and a half could have proceeded without extensive formal regimentation. But what these objections do indicate is that formalization is not the epistemic key to the problem of *a priori* knowledge. And this in turn suggests that we would do well to consider homely examples where our intuition registers a strong assent without (at least) any overt formal reasoning and to inquire into the rational basis of our confidence.

To take a shopworn example, consider the sentence "All bachelors are unmarried." Fluent speakers of English immediately recognize that this expresses a truth, and those uncorrupted by Quine recognize that it expresses a necessary truth. How do they do this? Not by deduction from evident first principles. Rather, so runs the traditional answer, one's recognition that this sentence expresses a necessary truth arises out of a knowledge of the *meanings* of the terms involved; and for just that reason one's knowledge of this truth is independent of one's knowledge of additional, empirical facts. That independence, of course, lies at the heart of *a priori* knowledge.

This traditional theory of analytic *a priori* knowledge can be found at least in embryonic form in the doctrine of intuition advanced by Descartes in his *Rules for the Direction of the Mind*, which was circulated among the Cartesians in manuscript after Descartes's death but not published until 1701, and in the very similar doctrine elaborated and defended by John Locke in book IV of the *Essay Concerning Human Understanding*. Locke's own formulation serves as a useful point of departure.

> It is the first act of the mind, when it has any sentiments or ideas at all, to perceive its ideas; and so far as it perceives them, to know each what it is, and thereby also to perceive their difference, and that one is not another. This is so absolutely necessary, that without it there could be no knowledge, no reasoning, no imagination, no distinct thoughts at all. By this the mind clearly and infallibly perceives each idea 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: and this it does without pains, labour, or deduction; but at first view, by its natural power of perception and distinction.[4]

To this faculty of the mind to know its own ideas and their "agreement or disagreement" Locke gives the name "intuition," and the knowledge that it furnishes he calls "intuitive knowledge." He contrasts it with demonstration, which is also (according to Locke) a source of certainty but is not an immediate perception of a relation between one's ideas. Rather, demonstration requires the interposition of steps of inference, each step being certified itself by an act of intuition.[5]

As a concept empiricist, Locke insists that the object of intuition is a relation between elements given in sensation, either mediately (that is, of complex ideas themselves derived from reflection upon ideas given in sensation) or immediately. Locke thereby takes on the challenge not only of generalizing from particular concepts, for which he invokes the doctrine of abstraction, but also of accounting for *all* of our concepts by means of such generalization. It is doubtful whether this latter goal can be achieved. But Locke's view has more resources than might appear and perhaps more than he appreciated himself. For the concepts of identity and distinctness themselves are presumably discernible *as* distinct by intuition, and they can

hardly be said to be sensuous. How those concepts themselves arise is a bit of a puzzle, but it creates more difficulties for Locke's empiricism than for his doctrine of intuition.

Locke's formulation of the doctrine of intuition contains key elements that suit it for an account of *a priori* knowledge of analytic truths. As the "first act of the mind," intuition is epistemically prior to other modes of knowledge; no regress looms. Though it may involve a comparison of concepts that were first given in experience, it does not depend upon some further experience that is the relation of those concepts: the relation is intrinsic and the perception of it infallible. And it promises to be the foundation for our knowledge of demonstrative truths as well.

That is not to say that intuition, as Locke describes it in this preliminary quotation, is equal to that task; it requires elaboration and refinement, some of which Locke himself undertook. The restriction to agreement or disagreement of concepts is too vague (a fact that has given commentators some grounds for frustration[6]) and, when made specific, too limited to do the full job. Intuition must yield not only the identity or distinctness of concepts but also such relations as containment and exclusion, and these will have to be at least generalized in order to allow intuition to underwrite our grasp of the cogency of reasoning by *modus ponens* or mathematical induction. But we need not have an exhaustive list of intuitable relations among concepts in order to underwrite at least some analytic truths; pretty clearly the four relations listed above will do for such truths as "black is not white," "green is a color," "nothing can be simultaneously red and green all over," and "dogs are not cats." It is important for epistemology that at-least-partial disjointness (failure of containment) be intuitable, and it is quite plausible that it is so in at least some cases. For suitably clear concepts we can intuitively grasp such epistemically significant truths as "success is not sufficient for rationality" and "rationality is not sufficient for success."

Locke speaks of the "evident lustre" of intuitive truths for the attentive mind, and the terminology of the "natural light" appears everywhere in Descartes. But it is vital not to confuse the epistemically significant act of discerning the relations of concepts with a phenomenological sense of great confidence in oneself or obviousness in the proposition.[7] Russell notes that "luminous obviousness, by itself, seems, as a matter of empirical fact, to be insufficient to insure truth; at least, if it is to be sufficient, it must be very carefully defined and limited."[8] Russell's point here seems to be that we would, to make it useful, have to find a way to parse out the metaphor in epistemic terms.

The natural way to explicate the metaphor, both interpretively in Locke and philosophically, is in terms of a relation of *direct acquaintance* that is *sui generis* and essentially relational.[9] One's knowledge of one's own concepts need not be mediated; there need not be any fallible steps by which one undertakes to determine what those concepts are. The importance of this point will emerge when we turn to criticisms of the Lockean view.

Furthermore, because we can know our concepts directly, we can also be directly acquainted with at least some of the relations between and among them, when we hold the related concepts clearly in thought at the same time. We have *a priori* justification for believing an analytic truth *p* when we see, by acquaintance with the relevant conceptual relations, that there is no alternative but for *p* to be true.

This conceptual analysis provides a plausible response to BonJour's twin challenges. Our acquaintance with the relations of our own concepts explains why analytic *a priori* knowledge of this sort is justified. And the required grasp of concepts and their relations cannot be called "mysterious" in any invidious sense. It is, indeed, quite easy to understand what this grasp is like, particularly in the case of very simple analytic truths.

The challenge to demystify analytic *a priori* knowledge is, we must remember, part of a *tu quoque*. BonJour's argument is not that some degree of mysteriousness is a bad thing *per se* but that analytic *a priori* knowledge as explicated by its adherents is no less "mysterious" than synthetic *a priori* knowledge. But BonJour sometimes seems to mean by "mysterious" merely fundamental or not further analyzable. This tendency arises when he discusses an example given by C. I. Lewis, who argues that it is an analytic truth that the relation "All X is Y" is transitive. Lewis says that one knows this fact by knowing what 'all' and 'is' mean and by understanding the syntax of expressions having the form, "All _____ is _____." Says Lewis, "One who understands meanings in English to that extent, will know that the relation so expressed is transitive."[10] BonJour is unimpressed, complaining:

> But merely listing the elements that would have to be grasped in order to understand the proposition provides no insight into *how* the proposition is known on the basis of those elements. Lewis seems to be saying merely that once those elements are understood, one can just see or grasp intuitively that the relation is transitive, a view that is, of course, entirely indiscernible from that of the rationalist.[11]

This is a very odd claim, since by definition synthetic *a priori* knowledge is *not* merely a matter of understanding meanings and their relations. Synthetic *a priori* statements are supposedly seen to be "true in all possible worlds" by an insight not reducible to conceptual grasp.[12] But it is precisely by means of discussing meanings (or concepts) and their relations that the analytic *a priorist* gives an account of justification. It is hard to see what more BonJour wants from Lewis by way of an explication as to "how" knowledge of meanings provides knowledge of the truth in question. Precisely because the truth is analytic, there is, on Lewis's view, nothing more involved in knowing it *a priori* besides a knowledge of meanings and the connections among them. The fact that conceptual analysis does not go on forever is in no way particularly damning to the advocate of analytic knowledge, nor

does it mean that analytic *a priori* knowledge is indistinguishable from synthetic.

The task of answering BonJour's criticisms is complicated by the fact that he does not deny entirely the existence of analytic *a priori* knowledge; indeed, he defends it against Quinean criticisms. But what he defends against Quine is an extremely attenuated version of analytic knowledge that he considers trivial or "reductive."[13] His discussion, therefore, gives a very strong impression that any *interesting a priori* truth is synthetic.

A traditional analysis of *a priori* knowledge does press meanings into service to define the notion of an analytic truth as one that is true in virtue of meanings alone and regardless of any matter of fact. In their heyday the positivists employed this notion of analyticity to resolve the conflict between judgment rationalism, according to which some non-analytic truths can be known *a priori*, and judgment empiricism, according to which the analytic and the *a priori* are co-extensive. Purported rationalist counterexamples to the equivalence of the analytic and the *a priori* might be admitted to be necessary but analyzed as analytic after all, demoted to the level of mere *a posteriori* knowledge, or simply analyzed away as emotive expressions.

Our purpose here is not to defend the sweeping positivist program for analyticity but rather to put the concept of analytic *a priori* knowledge to more modest use. At least some non-trivial necessary truths, among them the epistemic principles at the heart of our internalist epistemology, are analytic truths and can be grasped as such. This sort of *a priori* knowledge grants awareness of the truth and the necessity of propositions in virtue solely of acquaintance with concepts and with the relations among them; but it does more than that. To grasp a truth in this fashion is also to show decisively one's own justification to oneself. We are not claiming here that all necessary truth is analytic nor that all necessary truth must or can be grasped in this fashion. For our purposes it suffices that the concept of analyticity itself, explicated in terms of meanings, provides us with a plausible answer to the justification challenge posed by BonJour and can be applied to epistemic principles for purposes of stopping the metaregress.

But just here we must answer a chorus of objections: that meanings are external, that *a priori* knowledge is fallible, that concepts are not *things*, and that our theory would render it impossible to doubt conceptual truths. In the succeeding sections we take these up in turn.

Against semantic externalism

In an important sense our conception of analyticity is both radically internal and radically individualistic. Analyticity arises from the conceptual structure of the individual or, in the case of truths expressed verbally, from the meanings he attaches to the constituent terms. And those meanings, contrary to the dominant position in contemporary philosophy of language, we take to be wholly internal to the mind of the individual.

In his well-canvassed objections to semantic internalism, Hilary Putnam fixes our attention on the referent of a term.[14] The word 'water,' for example, refers to a certain sort of stuff with the chemical composition H_2O, a composition unknown up until the last two centuries. In everyday life we identify water by its manifest characteristics: it is colorless, odorless, tasteless, fluid at room temperature, potable and good for the garden. But in view of the well-entrenched chemical theory, we would be very reluctant to call something "water" if we had conclusive reason to believe that it were *not* H_2O. Which description, then, the chemical one or the common one, governs the *meaning* of the word?

Putnam suggests that the chemical description is central and that we cannot consistently deny that water is H_2O. Yet he acknowledges that the chemical structure of water was discovered *a posteriori* and could not have been discovered by any *a priori* means. Hence he holds the position, puzzling from a traditional standpoint, that some necessary truths are knowable only *a posteriori*. Worse, since in Putnam's view this truth is the key to the semantic question of the meaning of 'water,' the meaning turns out to be a function of factors "outside the head" of the user of the term.[15]

The conceptual analyst will approach this question in an entirely different fashion, wielding the conceivability criterion rather than scrutinizing the scientific record. We can conceive, after all, of the failure of the chemical description; it *might* turn out that water – that is, the stuff with the ordinary set of manifest characteristics – is not H_2O, current evidence notwithstanding. The point is not that anyone anticipates the overthrow of current chemical wisdom, but rather that the statement "water is not H_2O" is *intelligible* (i.e. the situation it describes is conceivable) and therefore possibly true. But in that case, we must be able to attach a meaning to the term independent of the chemical one – as in fact we do whenever we identify water by its manifest properties. Contrary to Putnam, therefore, the actual physical structure is not (or need not be) part of the *meaning* of the term 'water.'

One point in dispute between semantic internalists and semantic externalists is whether the conceivability test is a good one. Elsewhere we have responded to alleged counterexamples to the conceivability criterion of possibility.[16] But even if we are granted that point, semantic externalists could retort that Putnam's Twin Earth example establishes another conclusion incompatible with semantic internalism – that intension does not determine extension. This, in fact, is one of Putnam's major points in his seminal article, and he evidently thinks that it is devastating to traditional ideas of meanings as mental entities.[17] If two physically and psychologically identical counterparts, one on Earth and one on "Twin Earth," each point to a sample of the local colorless, odorless, tasteless, potable fluid and call it "water," they are in Putnam's example picking out something different; for on Twin Earth the fluid referred to has the chemical composition XYZ rather than H_2O. The identical makeup of the counterparts – including

their identical psychological states and hence identical conceptual meanings – does not, therefore, guarantee that their reference is to the same sort of stuff. If meanings determine reference and reference is determined by extra-mental factors, then are meanings not, or not altogether, "in the head" after all?

Just what this argument amounts to depends on how one takes the claim that meanings determine reference.[18] The claim that intension determines extension contains a core of truth when properly understood. Given one's *actual* environment, the true state of the real world one lives in, as a constant factor, the only way to vary the extension of a term is to vary the concept it expresses. (One of the recurring errors of semantic externalists like Putnam is the failure to recognize just how fine-grained are the distinctions that a radical semantic internalist can employ to individuate concepts.) If the people on both Earth and Twin Earth are using the everyday concept of water, without reference to chemical content, then the extension of the term for speakers on both planets will include both H_2O and XYZ, provided that both fluids exist in the world and support the manifest properties in question.

This sense of the phrase 'intension determines extension' helps us to analyze correctly another of Putnam's examples, the case of the elm tree and the beech tree. Putnam claims that, being ignorant of botany, he is unable to distinguish elms from beeches and that his concept of an elm is identical to his concept of a beech. If we take his word for it and imagine (he does not give details) that his concept for both words is something like "a tall tree having leaves with serrated edges," and if in fact both elms and beeches fulfill this description, then Putnam is simply *wrong* to insist, as he does, that the extension of the two terms is different in his idiolect.[19] Given such a broad concept attached to both terms, and given the actual structure of trees in the world, the extension of either term is the entire set of elms and beeches (and perhaps other species as well). A botanist will, by using a more precise concept, restrict the extension of each term in his own idiolect to a smaller set of actual trees.

However, in its strongest sense – that meaning alone, regardless of the external setup, suffices to determine uniquely what is referred to – the statement that "intension determines extension" or "intension determines reference" is clearly false. If Putnam's putative counterexamples have any force, they cast doubt on this unnecessarily strong principle of determination rather than on the internality of meanings. Meanings will constrain or broaden reference, to be sure; but the suggestion that they are the only thing that can constrain reference is more than a little strange. The environment also places some limits on what one can refer to, if only by the expedient of omitting to stock certain items. A person cannot refer to "my horse" if he does not have one; nor can he refer to water if none exists, nor to "the water in that cup" if the cup is empty.[20] And *pace* the causal theorists, a name like 'Sherlock Holmes' can carry meaning even though it has no referent.[21]

If by 'beech' Putnam means (*inter alia*) "a tree having the same underlying genetic structure as the type of tree I call an 'elm,' of which (ostending here) *this* is an instance," and vice versa, then the universe may not cooperate and both terms may fail of reference, since beeches and elms are not genetically identical. This is one of the pitfalls of trying to talk about the real world and incorporating bits of empirical knowledge into our concepts as we do so – the concepts may not be instantiated. This is not, of course, to say that we should never allow our concepts to be influenced by scientific knowledge. But neither does it show anything like the claim that "meanings ain't in the head." On the contrary, when we learn that a term we have used did not refer to anything real, it will usually be the case that our concept was a composite, and often, a composite of aspects not related to each other by necessity, as in the case of water. In that case, it is only by getting the various parts of the concept separated in our own minds that we will be able to tell which parts are instantiated and which are not. A person who uses the phrase 'my horse' doubtless intends many different things by the phrase. If he separates "the horse in front of me" from "the horse to which I have right of ownership," and both of these from "the cause of the sensations I am having at this moment," he will be able to see more clearly where his attempt at reference suffers from presupposition failure. The fact that intension neither absolutely determines extension nor guarantees reference thus demonstrates not the external nature of meaning but the tremendous importance of conceptual analysis.

Putnam considers this sort of approach unsatisfactory because it permits a situation in which two terms are both phonemically different and different in extension but have the same meaning. For example, consider the parallel universes where the two subjects have the same concepts for 'water' (for simplicity, we might take it that both of them are using the manifest property concept), but where one substance is XYZ and the other is H_2O. We can now vary the example just slightly by imagining that the words in the universes are also phonemically different, e.g. that the subject in one world calls the substance "quaxel." In that case, if we use 'meaning' to mean intension, the words 'quaxel' and 'water' have the same meaning in the two universes, although they have both different forms and different extensions.[22]

We can only say that we do not share Putnam's evident revulsion for this conclusion. The situation once again requires simply that we distinguish two senses of a word – in this case, intension and extension as senses of 'meaning.' Since the former is, from our point of view, by far the more interesting of these, we are strongly inclined to use 'meaning' in the first sense. But regardless of usage, when the two are distinguished there is nothing remotely paradoxical or counterintuitive about saying that the words in the two universes have the same meaning (in the sense of intension) while having both different extensions and different forms. There is no convincing argument here for Putnam's contention that we should give up on treating meanings as mental entities in any sense whatsoever.

If Putnam's arguments for the radical externalization of meanings are not compelling, as we have argued they are not, then we are entitled to investigate the prospects for a traditional notion of conceptual truth in terms of the relations among intensional meanings, letting reference fall where it may. Of course this would be problematic if our theory *prevented* reference (even indirect reference) to objects with which we are not directly acquainted. But this is precisely the difficulty solved by Russell's theory of descriptions.[23]

Phenomenology and fallibilism

Knowledge, it is widely conceded, involves belief. From this conceptual truism Alvin Plantinga mounts an attack on the notion that *a priori* knowledge is an infallible source of beliefs. For S to know *a priori* that A, he writes, is first of all "for S to *believe* A and *believe* it *a priori*."[24] Plantinga then sets out on a quest for the analysis of *believing a priori*, which he claims is a special sort of believing that occurs when the only sensuous experience S must have to believe that A is whatever sensuous experience is required for S to grasp it.

But the sensuous imagery that typically accompanies a belief of this sort is not, in his view, the most important aspect of *a priori* believing, for imagery may vary from one believer to another. Rather, what is distinctive about *a priori* believing is the accompanying *feeling of rightness* about the belief, a sort of phenomenology that is not (or not wholly) sensuous.

> These beliefs seem wholly obvious; I find myself utterly convinced. They have about them, furthermore, the peculiar feel that *a priori* beliefs have – that feel that somehow they just couldn't possibly be false.[25]

To *see* the truth of a proposition *p*, according to Plantinga, is:

> (1) to form the belief that *p* is true and indeed necessarily true (when it *is* necessarily true, of course), (2) to form this belief immediately, rather than as a conclusion from other beliefs, (3) to form it not merely on the basis of memory or testimony (although what someone tells you can certainly get you to see the truth of the belief in question), and (4) to form this belief with that peculiar sort of phenomenology with which we are well acquainted, but which I can't describe in any other way than as the phenomenology that goes with seeing that a proposition is true.[26]

What is it, then, to *believe* something *a priori*? Simply this: that one believes it in a fashion that meets the conditions for *seeing* its truth (or that one sees that it follows from something else that one sees to be true), except that the truth condition may or may not be met. The phenomenology, which (if we

take Plantinga's statements at face value) amounts to a strong sense of psychological compulsion, can of course occur in the absence of the truth of the proposition.

> [S]uch a feel could be misleading. A false belief, obviously enough, could have that sort of feel for me: I could be mad, or a victim of an Alpha Centaurian cognitive scientist, or a brain in a vat, or a victim of a Cartesian evil demon. Indeed, it was in this very context of *a priori* knowledge that Descartes turned to that evil demon scenario. So obviously I can be wrong, even when it seems *a priori* for all the world that the belief is true. As a matter of fact, this isn't merely an abstract possibility: some propositions that have that *a priori* feel about them *are* false, as is shown by certain versions of the Russell paradox ... [27]

Hence, on Plantinga's account, we can believe *a priori* but falsely; and to think otherwise is to be misled by phenomenology. Conversely, *a priori* knowledge is just "believing *a priori*" together with the truth of the proposition believed. This analysis makes *a priori* knowledge very much like "scientific knowledge" or "sensory knowledge." It is only statistically correlated with truth rather than being *by its very nature* a type of access to truth. "*A priori* believing," on Plantinga's view, is the operation of the faculty of reason. And the use of this faculty is an *attempt* to discover truth without the use of sensory input. If one happens to have *a priori* knowledge, this is, as in the case of other fallible belief sources, merely a matter of successful *a priori* believing.

As an attempt to capture the traditional notion of *a priori* knowledge, this analysis is fatally flawed. The first and fundamental error occurs when Plantinga assumes without argument that there is a species of belief properly termed '*a priori* believing,' a conclusion to which he is not entitled merely because there is a species of knowledge to which the term '*a priori*' applies. If this is not a fallacy of division, what is it?

Plantinga seems to misunderstand the nature of direct acquaintance as put forward by many traditional foundationalists.[28] For acquaintance is fundamentally relational; and an account of *a priori* knowledge in terms of acquaintance will involve, ineliminably, grasp of the *content* of the proposition believed. Such a grasp cannot be reduced to *frissons* of delight that may accompany it. Treating *a priori* knowledge as true belief with – on the internal side, at any rate – nothing but epistemically irrelevant phenomenology, he has left the advocate of *this* theory of "*a priori* knowledge" without the ability to give any internally accessible answer to the question of how such knowledge has its positive epistemic status. It may be a boon to an externalist to cast the issue in this form; but Plantinga's conception of the *a priori* is a straw man.

Plantinga's own epistemological position, as BonJour points out, does not give us *any* account of the connection between the character of experience

and the content of belief.[29] We would carry BonJour's critique one step further: Plantinga's account of *a priori* knowledge does not merely render mysterious the connection between, on the one hand, what we mean by 'red' and 'color,' and on the other our *a priori* knowledge of the truth that all red things are colored; it positively *severs* the epistemic connection between the structure and interrelations of our concepts and our knowledge of the analytic truths we express using those concepts.

In the discussion thus far we have taken at face value Plantinga's assertion that he is talking about a mere "feeling of rightness" when one contemplates a proposition. Another interpretation is possible. Perhaps when Plantinga speaks of "that peculiar sort of phenomenology with which we are well acquainted . . . that goes with seeing that a proposition is true" he is indeed talking about what we have spoken of as being acquainted with the truth of an *a priori* proposition or seeing clearly and distinctly the relations of one's concepts.

But if that is what Plantinga means, then perhaps his argument for *a priori* fallibility is more pressing than we have been acknowledging. Suppose we are granted for the sake of the argument that there are analytic truths knowable on the basis of an intuitive grasp of the relations among their constituent concepts. Some people, let us say, do in fact believe such truths in this fashion and are utterly convinced of the truth, perhaps even of the necessity, of what they believe. Still, others believe falsehoods with no less conviction – they go wrong in matters that seem "as clear as sunlight," to use a Cartesian phrase. If this is true, how can any one of us know whether his apparently indefectible grasp of a putative conceptual truth is of the former or the latter kind? How can he know whether he stands, epistemically, with the sheep or the goats? Ernest Sosa voices this very objection:

> After all, it is not as though we are infallible about everything in these realms. It is not as though here we could never make a mistake. And even if we take ourselves to be infallible and incorrigible about certain of our beliefs at least in these realms, how would we know which of them are truly so protected against error?[30]

Similarly, Plantinga levels this challenge at Richard Fumerton:

> Where can [the internalist] go for assurance that what seems to him self-evident – 2 + 1 = 3, for example, or the thought that *modus ponens* is a valid form of inference – are really so? . . . Following Fumerton, he may perhaps say that at certain points he has a direct acquaintance with reality, a direct intuition of the fact involved. But . . . even in these very situations he can be mistaken. So all he can do, really, is assume that indeed he is sometimes directly acquainted with some states of affairs, and that, *usually*, at least, when it seems to him that he is directly acquainted with some state of affairs, he really is.[31]

But there is in all of this the unspoken and prejudicial assumption that someone who grasps a necessary truth in the indicated fashion does not *thereby* know where he stands. To put the same point in a different way, Plantinga and Sosa are assuming that real *a priori* grasping is epistemically indistinguishable from feeling very confident about a proposition (while thinking hard about it) when one is in fact making a mistake. This supposedly follows from the fact that people have thought they had true *a priori* insight in areas such as mathematics or logic when in fact they were wrong.

But this does not follow at all. To understand this point, we must recognize a fundamental asymmetry. Both those who genuinely see a truth *a priori* and those who err while contemplating a proposition may think that they are having an experience of genuine *a priori* grasping. But their experiences are in fact not identical in type. Once again, we are asserting a radical difference between *a priori* knowledge and sensory knowledge. Here the situation is not like that in which one could be having all of one's sensory experiences, indistinguishable from the experiences in a veridical case, yet be the victim of a deceiver. Rather, our position is that genuine *a priori* grasping is a unique form of epistemic experience that grants not only assurance of the truth of the proposition in question but also of one's own indubitable justification in accepting it.

If one *could* (also clearly and distinctly) "compare" the experience of a genuine grasping of some truth – truly seeing it clearly and distinctly – with an only partially successful attempt to think about it as clearly as one can, one would indeed be able to see the difference between these two epistemic situations. But the person in the second situation is, *ipso facto*, *not* seeing the proposition clearly and distinctly, and so of course he is not in a position to make such a comparison. It does not follow that he will always be over-confident, will always think himself to be having a genuine *a priori* grasp when he does not. He *may* be able to recognize the strain or fuzziness, the lack of clarity and distinctness in his understanding of his concepts. But it is not a necessary truth that he will be able to do so in all cases. And this fact in no way impugns either the infallibility or the metaregress-stopping nature of *true a priori* grasping. Only a perverse sort of verificationism could require that A cannot have intuitive knowledge in situation S_1 unless he has a method for determining, in all situations S_n where he does *not* have intuitive knowledge, that he does not. One may sometimes mistake the call of the jackal for the roar of the lion, but it does not follow that one can mistake the roar of the lion for the call of the jackal.

It is true, as Plantinga points out, that Descartes raises at least the specter of fallibility in the context of discussing simple necessary truths in the first Meditation:

> [S]ince I sometimes believe that others go astray in cases where they think they have the most perfect knowledge, may I not similarly go

> wrong every time I add two and three or count the sides of a square, or in some even simpler matter, if that is imaginable?[32]

But curiously, just two Meditations later we find Descartes in a most unskeptical mood about these matters:

> [W]henever 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 utterly clearly with my mind's eye. Yet when I turn to the things themselves which I think I perceive very clearly, I am so convinced by them that I spontaneously declare: let whoever can do so deceive me, he will never bring it about . . . that two and three added together are more or less than five, or anything of this kind in which I see a manifest contradiction.[33]

What is significant in these two passages is that the doubts are always raised indirectly, as a possibility of failure in a situation under a description; the certainty comes with the actual focus on the proposition in question. Thus, Descartes raises doubts about whether he might go wrong when he tries to "add two and three," the operation not being carried out but rather alluded to; but when he actually turns his attention to the proposition that two and three added together come to five, all doubts vanish. The skeptical problem is abstract and arises only when we are not paying attention to the propositions themselves, a point he makes explicitly when he replies to his critics. The parallels between Descartes and Locke are notable, both in substance and in imagery:

> [I]f there is any certainty to be had, the only alternative is that it occurs in the clear perceptions of the intellect and nowhere else.
>
> Now some of these perceptions are so transparently clear and at the same time so simple that we cannot ever think of them without believing them to be true. The fact that I exist so long as I am thinking, or that what is done cannot be undone, are examples of truths in respect of which we manifestly possess this kind of certainty. For we cannot doubt them unless we think of them; but we cannot think of them without at the same time believing that they are true, as was supposed. Hence we cannot doubt them without at the same time believing they are true; that is, we can never doubt them.[34]

The mere fact that we cannot doubt such propositions does not guarantee that we are right about them. Descartes is aware of this natural objection, and in response he argues that it misses the point:

> It is no objection to this to say that we have often seen people 'turn out to have been deceived in matters where they thought their

> knowledge was as clear as the sunlight'. For we have never seen, indeed no one could possibly see, this happening to those who have relied solely on the intellect in their quest for clarity in their perceptions; we have seen it happen only to those who tried to derive such clarity from the senses or from some false preconceived opinion.[35]

Here we can find a plausible explanation of Descartes's frustration with his critics over *a priori* knowledge. Yes, of course, humans often err in their judgments, even when they feel firmly convinced of their rightness; but what does that prove about one's grasp of the relation of the concepts involved in the belief that all red things are colored or that the law of non-contradiction is true? As for the scope of skeptical doubt, the notion of a powerful deceiver seems to sweep all before it only when we are not actually contemplating a truth so simple and clear that we cannot understand what it means without seeing it to be true.

What, then, becomes of the skeptic? We have argued that metaregress-stopping *a priori* knowledge is possible and that its possibility is not in any way undermined by the fact that people make mistakes, sometimes with a strong sense of conviction, in non-empirical contexts. It does not follow, of course, that any given person, or anyone at all, *has a priori* knowledge as we conceive it. And if we as internalists are concerned with answering the skeptic with respect not only to object level beliefs but also to epistemic principles, are we not guilty of letting ourselves off the hook if, in the final analysis, we only gesture in the direction of a *sui generis* infallible grasp?

But although confrontation with a skeptic who questions a (putatively analytic) epistemic principle may present a difficult dialectical situation, a failure to convince him does not in fact point to a failure of internalism to meet its own norms. It is precisely because concepts are internal to a given knower that one person may not be able, so to speak, to "give" his own concept and his clear and distinct analysis of it to someone else. This is why philosophy proceeds by examples and counterexamples, by intuition pumps and explications, rather than by pointing to empirical facts. In such dialectical situations we are attempting to convey to the other person the ways in which our own concepts intersect and the ways in which his conclusions are (if we disagree) incorrect as analyses of these concepts. Just because justification is radically internal, this is the best we can do. If our interlocutor is not using the same concept we are using, or if one or both of us are not seeing it with sufficient clarity, there is no more to be done dialectically. The definition of *showing decisively* in Chapter 4 involves the requirement for stopping the metaregress that it be impossible rationally to doubt the justification of a belief at a given level. If A is genuinely grasping an epistemic truth and B is not, it is indeed not possible for A *rationally* to doubt either the truth or his own justified status; but B is not seeing that proposition in itself with perfect rationality. To demand the adducing of some *separate*

criterion – separate from the nature of the grasping as directly accessible to the knower – as part of an *argument* that the experience is one of genuine grasping rather than an *Ersatz*, is to misconceive the nature of *a priori* insight and to rule out metafoundationalism at the outset. Those readers who have had an experience of direct, clear and distinct *a priori* grasping will, we hope, recognize the description and see the potential of such a grasp for grounding epistemology and escaping the metaregress.

Plantinga has one additional arrow in his quiver. If the traditional view is correct, then the epistemic value of *a priori* knowledge is always the same: genuine knowledge *a priori* yields certainty. But does not intuitive justification come in degrees? Confronted with the Russell paradox, he writes, no one

> proposes to rest in the conclusion that there is a property that both does and does not exemplify itself; nor does anyone propose to solve the problem by giving up *modus ponens*, even though the latter is essentially involved in the argument. Instead, we pick some premise that seems less certain, one that has less warrant.[36]

Here again, the criticism hits a target of Plantinga's own making but misses the traditional view he takes himself to be criticizing. The clue lies in his gloss in the final line: " . . . that seems less certain, . . . that has less warrant." The notion of warrant is idiosyncratic to Plantinga, and he is free if he likes to define it as varying in intensity according to how certain things seem – presumably, how they seem to a properly functioning individual. But though subjective confidence comes in degrees, such confidence may be well or poorly placed. No particular degree of it is a reliable index of *a priori* justification invoked by the traditional epistemologist.

It is true that our degree of rational confidence even in mathematical or logical propositions may vary as we have better or worse evidence for them. But rather than disproving the existence of traditional *a priori* knowledge, this reveals the varied and interesting nature of the rational arguments we can make *a posteriori*. In fact, Frege's response to the Russell paradox – Plantinga's own example – supports this point. When Frege attempted to found mathematics on self-evident principles of pure logic in the *Grundgesetze*, he assumed a naive principle of set formation that permitted, in conjunction with his other principles, the derivation of a now famous contradiction. The young Bertrand Russell, writing to Frege in 1902, pointed this out, whereupon Frege commented that the principle of set formation in question had never seemed as strong a principle as the others.

> I have never disguised from myself its lack of the self-evidence that belongs to the other axioms and that must properly be demanded of a logical law.[37]

He attempted to rectify the problem in various ways without notable success, and at the end of his life he abandoned the attempt to ground arithmetic in logic and attempted instead to derive it from geometry.

Susan Haack takes Frege's famous slip to support fallibilism by showing that "there can be serious doubt which statements are, and which are not, *really* self-evident," concluding, "Because people disagree about what is self-evident, self-evidence is, again, useless as a sign of certainty."[38] But this is a very strange moral to draw, as is Plantinga's conclusion that *a priori* justification comes in degrees. Surely Frege's own statement, taken at face value, indicates not that he was ever in serious doubt as to which statements were self-evident, but that he knew quite well all along that the axiom of abstraction was *not* self-evident and used it for some other reason. Indeed, Frege's case seems to be one in which the subject believing a putative logical truth realizes that, even if it is true, he is not grasping it clearly and distinctly. Hence, on our account, even if he is right, he does not have *a priori* knowledge; some use of *a posteriori* reasoning is involved.

It is unquestionably possible to have an *a posteriori* belief *about* the truth of a mathematical proposition. To take a very simple example, a student may believe that a particular formula is a mathematical theorem because he has reason to trust the textbook in which it appears. Even if he is correct and the formula does represent a necessary truth, he does not know it to be true *a priori*. Even when we are carefully contemplating particular true logical or mathematical propositions, if we have (perhaps because they are complex) only an indistinct grasp of them or merely a strong hunch that they are right, we cannot know them *a priori*. Rather, we are introducing an inductive or other *a posteriori* element involving some such premise as, "Usually when something looks right as this does, when I am able to check it out clearly and distinctly, it has turned out to be true." Whether or not we are right in such cases, we are not, as Descartes would put it, "relying solely on the intellect." Frege's adoption of the axiom of abstraction may have been a case in point; the axiom is a generalization that has some true instances, and generalizing is a natural but fallible method of reasoning. It is because we can reason about necessary truths in these indirect ways that we can have varying degrees of rational confidence concerning them, not because real *a priori* knowledge is itself only fallible.

Ontological worries

Much of what we have been saying seems to indicate that concepts or meanings are things. Any such idea is distasteful to philosophers of Putnam's school; Putnam even puts the term 'mental' in scare quotes when discussing mental states and dismisses the idea of entirely self-contained psychological states as "methodological solipsism."[39]

It is fairly obvious that Putnam dismisses the possibility of real mental states on the basis of an inclination to materialism and that Quine dismisses

meanings because of a strong commitment to behaviorism. But even though BonJour has no stake in either materialism or behaviorism, he is also uneasy with the idea that concepts or meanings are entities. He describes meanings (for purposes of defending an attenuated notion of analyticity) as "*whatever* it is in virtue of which an expression is meaningful in the way that it is" and notes with satisfaction that this avoids a "view of meanings as 'queer entities' whose existence might be metaphysically problematic."[40] Elsewhere he states the concern at greater length.

> Clearly the main difficulty ... is to get clearer about what sort of thing a *concept* is supposed to be. While it is clear enough that concepts are at least roughly the philosophical descendants of the *ideas* invoked by earlier philosophers like Locke, and also that talk of concepts (or ideas or notions) often seems virtually unavoidable in philosophical discourse, none of that helps in any very immediate way to clarify exactly what such talk is about.[41]

The issue is a slippery one, since there is not general agreement as to the sort of account of the nature of concepts one ought to give. A *definition* of "S possesses the concept of X" can be all too easily conflated with a *criterion* for "S possesses the concept of X." This seems to be the direction BonJour is headed, for in an attempt to clarify what "concept" talk is about he reduces possession of a concept to possession of certain abilities:

> [T]he possession of a concept of X by a person is to be identified with that person's having a certain cluster of intellectual abilities: the ability to think of X's, to classify things as X's, and, in some cases at least, to recognize X's in appropriate circumstances.[42]

But none of this, he goes on to say, "makes it very clear how a concept can be an object of knowledge in a way that makes knowledge of concepts an alternative to knowledge of the world."[43]

On this last point he is unquestionably correct, but the problem seems to lie not with the notion of conceptual truth or knowledge of concepts but rather with the assimilation of *possession of a concept to possession of certain intellectual abilities*. The latter, as BonJour defines them, are all directed at the object of thought – zebras, say – which we can think of, sort, and sometimes recognize. And there is no obvious reason that possession of these abilities requires that one possess a concept itself in any interesting sense. Indeed, the concept has more or less vanished on this account. And therein lies the problem. Reducing concepts to mere capacities makes it difficult to give an account of our ability to recognize those relations of necessity that BonJour himself wishes to incorporate in his neo-rationalism. If classification and recognition are to be more than the outputs of black-box mechanisms, we need some indication of the mental processes by which we

perform these activities. This is why possession of a concept cannot be identified with possession of abilities: the latter are inscrutable unless we can give some sense to the former.

What follows, then, is a list of some necessary conditions for something to be a concept, in the sense in which we are using the term. We do not claim that these conditions constitute a complete explication of the concept of concepts, but they will go some distance to indicate what sort of thing a concept is. Furthermore, it should be clear from these conditions that concepts do not present any special ontological problems beyond those presented to the materialist by *any* non-material entities. Since it is beyond the scope of this book to defend dualism against its various rivals, we shall have to let the physicalists think what they will and be satisfied with pointing out that concepts as we define them ought to be unproblematic for anyone who allows for the existence of minds.

First, a concept as we have been using the term here is a mental entity; it cannot exist if no mind exists. We are therefore not embarking on Frege's project to make meanings "public" by making them *abstracta* having a mind-independent status like that usually attributed to propositions. That project has its own interest, but a concept or meaning in that sense would be different from a concept in the sense we are here exploring. A concept in this sense, very much like an experience, is an aspect, element, or state of a mind, although not an essential one. That is, it is not necessary for a mind to be a mind that it have any particular concepts, or indeed any concepts at all, as an unconscious mind that has never possessed any concepts is conceivable.

If concepts are mental entities, this removes one possible source of ontological worry. It is perhaps understandable that *abstracta* like propositions should seem "metaphysically problematic," even to someone not committed to materialism,[44] but if concepts are *someone's* concepts and hence aspects of his mind, they are no more metaphysically problematic than experiences or thoughts.[45]

Second, a concept may be an object of direct acquaintance. As a person is directly acquainted with his own experiences, so also he has direct access and can refer directly to his own concepts.[46] Hence, concepts admit in principle of analysis entirely from one's armchair.

Third, a concept may be either simple or complex, and if it is complex, a given subject may not be able to advert to all its parts clearly and distinctly at the same time.

Fourth, the contemplation of a concept need involve neither visual imagery nor linguistic expression (even private linguistic expression), although it can be accompanied by either of these in some particular case. A subject can use a concept at a particular time as the meaning (intension) of a term.

Fifth, a concept need not be complete in a subject's mind, nor does a subject need to understand all of its entailments clearly and distinctly, in order for him to be acquainted directly with some portion of the concept.

The first point may raise a question regarding the use of mental entities as the constituents of conceptual (analytic) truths. Those attracted to a correspondence theory of truth may wonder whether treating concepts as mental entities compromises the position that analytic truths are not "about the world." Are not mental entities "in the world"? If concepts are mental and if conceptual truths have their truth value in virtue of the relations of concepts, would analytic claims be false – through a failure of correspondence – in a world without subjects?

But whether we treat concepts as abstract or as mental entities, we should insist that in one important sense analytic claims are not *about* anything at all. That is to say, they do not describe any state of affairs. To say that analytic truths are true in virtue of conceptual relations is not to say that they are *about* concepts, so that in the absence of any concepts they would all be false. S may know by means of an acquaintance with his own concepts that "red is a color" is necessarily true, but the statement "red is a color" is not the same as the (obviously synthetic) statement, "S's concepts are related in such a way that his concept of redness includes his concept of color." The latter would be true in virtue of the instantiation of the state of affairs it describes. It is contingent because it makes reference to S and his concepts, and these do not necessarily exist. The former is, so to speak, "all structure" and does not owe its truth-value to any actual state of affairs. The correspondence theory of truth, then, can be regarded as not applying to analytic truths; or, if we prefer, we may say that analytic truths cannot fail to correspond to reality, since there is no state of affairs that could make it the case that they are false.

Uncertainty, conceptual learning, and analyticity

In arguing that some non-trivial necessary truths are analytic, we encounter a different type of objection from any discussed thus far. If a necessary truth is analytic, if it is true in virtue of the relations of the concepts it uses, how is it *possible* for anyone to be initially uncertain about it? Furthermore, how is it possible for a subject to learn such a truth when he did not previously believe it, or perhaps even thought it to be false? If these truths are analytic and true in virtue of the relations of his *own concepts*, should not their truth be immediately obvious? And yet it is generally the case that, the more interesting a necessary truth is, the less obvious it is, at least for those of us mortals who do not have all our concepts perfectly regimented and clearly and distinctly before our minds at all times. Will we then be forced to concede that non-trivial necessary truths are synthetic after all?

To answer this family of questions, we must distinguish among three different possible sources of initial uncertainty about the truth-value of a putatively analytic necessary proposition. First, we might be initially uncertain about a claim because it involves concepts that are intrinsically fuzzy and because the claim concerns a relation of concepts at their "fuzzy

edges," so to speak. Second, we might be unsure about how one concept relates to another because the concepts involved, though not intrinsically fuzzy, have been only partially analyzed. Third, and relatedly, we might be initially unsure about, and hence able to learn, an analytic necessary truth by way of coming to see a relation of conceptual entailment.

One of the risks we run in advocating an appeal to concepts, particularly when we are dealing with intrinsically fuzzy concepts, is that the picture of *a priori* knowledge will be unrealistically rigid. It would be a brave speaker of the language who claimed to have an absolute answer for every question that might be asked about redness. There is, for example, a certain amount of arbitrariness involved in deciding whether or not to accede to the claim, "Nothing red all over is dark pink all over at the same time." After all, dark pink may be thought of as at least very close to red. One immediately wants to ask, "How dark is this dark pink supposed to be?" Similarly, if one encounters the statement, "No one who shows more than a square centimeter of skin on his chin has a beard," one is more inclined to read this as part of a ridiculous government regulation (perhaps under some extreme regime that puts high value on having a beard), establishing by fiat what counts as a beard, than as a necessary truth.

In these cases, the concepts themselves do not immediately show the statements to be true or false, and indeed the very indifference with which we greet the statements indicates that it truly is a matter of one's decision regarding word usage, unconstrained by any core meaning of the terms involved, whether to accede to them or not. But, someone might wonder, if we have an incomplete concept of what it is for an object to be red, how can we be, in virtue of the concept alone, certain of *any* claim in which that concept figures?

There is less to this objection than meets the eye. If one's concept is in some respects indefinite – as many of our concepts appear to be – this does not preclude the possibility that it quite definitely and unambiguously excludes other concepts. What is most interesting about the otherwise rather trivial subject of fuzzy concepts is the fact that even fuzzy concepts can be used as part of *unquestionably* necessary truths. If, for example, someone says, "The color red is not a musical note," we shall agree with this instantly as an obviously necessary truth, despite the fact that we are uncertain about the relation between redness and dark-pinkness. For whatever exact shades are included in the color red, we can tell absolutely that our concept of it excludes the concept of a musical note. The comparison of the concept of redness to the concept of a musical note thus concerns the clear center of our concept of redness, not its fuzzy edges. Questions about shade arise only *because* red is a color. Uncertainty about some applications of a concept does not amount to uncertainty about the concept in its entirety; therefore, even statements using "fuzzy" concepts can be true in virtue of meaning.

A more tricky subject is that of incomplete concepts that are not *intrinsically* fuzzy. These are especially prevalent in difficult subjects like probability

theory, epistemology, and mathematics. There is no bare decision, no mere linguistic convention, involved in the properties of prime numbers, as there is in the exact shades we are willing to call "red." Yet no one is certain about the truth-value of the Goldbach conjecture.

In these areas, a self-aware subject experiencing uncertainty about what he knows to be (if true at all) a necessary truth realizes that his uncertainty does not arise from his not having yet gotten around to deciding whether to call something by a certain word. Rather, he realizes that he does not have in clear focus in his mind all at once all of the parts of a concept that, in some of its aspects, he already has, or else that he does not yet see clearly all of the things that concept entails.

But here as well we can see both how uncertainty arises and that it does not preclude certainty about some analytic truths involving incomplete concepts. A subject may not be certain exactly where frequencies fit into his concept of probability, but he can still be absolutely certain that "x is probable" does not include "x is square." Indeed, he may very well be able to tell with certainty that the kind of thing that can be probable is not the kind of thing that can be square, and that in virtue purely of the concepts of probability and squareness. Clearly, the relation of the concept of frequency to that of probability is more complicated and delicate than the relation of the concept of squareness to that of probability. We realize that frequencies may have *something* to do with probability, since a knowledge of frequencies often leads a rational subject to make particular judgements about probability (or to believe particular propositions with a particular degree of confidence). It is only when we are told that probability just *is* frequency that we begin to wonder if this statement is too strong and are forced to engage in careful conceptual analysis in order to evaluate it. None of this, however, tends to show that the truths involved are something other than conceptual in nature.

To take a more difficult example, consider a subject with a broadly internalist concept of justification that he has not worked out in all of its details. For example, he may be sure that internal access is essential to justification but have no theory as to how this intuition applies to foundational beliefs; or he may not have thought about the implications of his internalism for philosophy of language or for various aspects of metaepistemology. Confronted by the Gettier problem as an argument for externalism, he may well be initially puzzled and uncertain about how the Gettier problem is relevant to his theory of knowledge. But this does not indicate that his epistemic beliefs are synthetic. For even a partial analysis of his concept of justification will show him clearly, and by conceptual analysis alone, that it cannot be true that a person is justified in the absence of *any* internally accessible epistemically relevant factors (such as either evidence or, for foundations, the nature of the belief itself that makes it self-justifying). He may wonder initially whether he needs to *add* some requirement to his concept of justification to deal with the Gettier problem, but he can be conceptually certain even before settling that question that a version of externalism that

makes "justification" entirely a function of factors that are (or might as well be) inaccessible to the subject, is incorrect. He can be confident that *mere* production by a reliable mechanism, for example, would simply not be justification as he conceives it, since he has realized that his concept of justification contains an irreducible internal component. This confidence may itself be a result of earlier conceptual analysis, but even though the analysis is incomplete, it can yield certainty about interesting and difficult analytic truths involving the concept.

Conceptual entailment presents a special challenge for the advocate of non-trivial analytic truth. To discuss the subject clearly, we must first give, as examples, some stipulative definitions of relations of conceptual entailment. Strictly speaking, a concept is not a truth-bearer and hence cannot entail anything. So,

> the concept of A *conceptually includes* the concept of B just in case, in virtue solely of the nature of the concepts themselves, it is a necessary truth that (x) (Ax→Bx). The concept of A *conceptually excludes* (or *analytically excludes*) the concept of B just in case, in virtue solely of the nature of the concepts themselves, it is a necessary truth that (x) (Ax→−Bx). Both conceptual inclusion and conceptual exclusion are relations of *conceptual entailment*.

The notion that there is such a thing as conceptual exclusion enables us to answer a criticism leveled by BonJour, who challenges C. I. Lewis's explication of truth in virtue of meaning on the grounds that it does not seem correct to say that all entailments of a predicate are "contained in" the meaning of the original predicate. BonJour argues that Lewis is left without any non-mysterious way of explaining the justification of various (from Lewis's perspective, analytic) truths, because these particular truths do not seem to involve any containment relation.

> The underlying point here is that the initial "criterion in mind" in terms of which one understands a term or proposition plainly does not include in any straightforward way all that is in fact entailed or necessitated by it. Much of the epistemological problem of *a priori* knowledge is precisely the problem of how to justify the transition from our ordinary grasp of meaning or content to the further entailed consequences. And thus to simply include those consequences in the intensional meaning, as Lewis does, offers no real epistemological gain: the original problem simply recurs as the problem of how to justifiably make the transition from the narrower ordinary meaning to the full intensional meaning.[47]

The proposition in question in the immediate context is "A thing that is red all over is not green all over." BonJour points out that the concept of

not being green all over does not seem to be in any obvious sense "contained in" the concept of being red all over. This is clearly true if one means by 'contained in' that the idea of not being green all over springs to mind whenever one thinks of being red all over. A similar, and even more obvious, example concerns numbers. "Exactly five" conceptually excludes, as conceptual exclusion is defined above, an infinity of other numbers both below and above five on the number line, yet we fairly obviously do not have an infinite number of thoughts – "Not six, not seven, not eight," etc. – in our minds whenever we think of exactly five. BonJour therefore concludes that truths of this sort should be understood not as conceptual truths but rather as synthetic *a priori* truths, since our knowledge of them apparently does not arise from our knowledge of what is included in our intensional meanings.

The metaphor of conceptual containment on which Locke leans is at best suitable for expounding the sorts of relations found in the traditional forms that make up syllogisms. The concept "animal," for example, is contained in the concept "man," and the concept "living thing" is contained in the concept "animal," and so these may be chained together by an intuition of the relevant containment relations. But containment by itself is manifestly inadequate as an account of deductive inference as it extends beyond the syllogism. Even in elementary modern logic we need to connect not just concepts but entire propositions. There is, at first sight, no helpful sense in which "John is a bachelor" either contains or is contained in "Either John is a bachelor or Elaine has read Frege's critique of psychologism," though the latter is derivable from the former.[48]

There is, however, no reason to restrict our notion of analytic truths to those involving containment; as we noted earlier, other relations can be intuitable as well. Borrowing some helpful terminology from John Pollock, we can say that there are *intuitions of implication* of which conceptual containment, exclusion, and so forth are special cases.[49]

When we speak of intuitions of implication we come close to an idea that is itself conceptually primitive, where we can do little more in the way of explication than repeat the idea. A person who has had the experience of seeing something to be true and necessary in virtue of its conceptual content ought to recognize the notion immediately. One thing, however, that we can emphasize again is that knowledge of a conceptual truth via an intuition of implication does not require an accurate understanding of facts about states of affairs in the world. To know that a thing that is red all over is not green all over, a subject need not have a correct theory of the propagation and refraction of light (for example), nor does he need to know that there are or ever have been any red or green objects.

BonJour's rejection of (at least significant) truth in virtue of meaning has as a correlate an insistence that *a priori* truths are indeed about the world. But this proposal involves a rather surprising redefinition of what it means for a claim to be "about the world." He states that even if there have never

been any objects answering to the description in a necessary truth, the truth is still "about the world" because it supports counterfactual statements about the world. So, for example, he takes the claim that a triangle must have three sides to be "about the world" because it supports the counterfactual truth that any really existent object that was a triangle would have three sides.[50] But this, surely, is not what is usually intended when one speaks of a claim as being "about the world." We can freely grant that analytic statements support counterfactuals in this way without thereby impugning their purely conceptual nature. Analytic statements are *true* even in a world in which the concepts they use are not instantiated, and their truth-value, as we have already argued, does not depend upon their correspondence to states of affairs in the world.[51]

In the case of comparing "red all over" to "green all over," we can see that the conceptual exclusion holds in virtue of the fact that, if something is a certain color at a certain location, this by definition means that this color fills the space at that location. "Red at n" is thus conceptually incompatible with "some-color-other-than-red at n." If the colors are truly different – and the concept of green is clearly incompatible with what we have called above the "core" of the concept of red, fuzzy edges notwithstanding – then they cannot both occupy the same bit of space, by the basic conceptual understanding of what it means to say that something is one color rather than another. This analysis gives us a further understanding of how "red all over" excludes "green all over" solely in virtue of the concepts involved and in a non-mysterious fashion. The example also illustrates the fact that understanding a new conceptual entailment can lead to a clearer understanding of the concepts involved, and this even when one concept is not psychologically "contained in" the other with which it is compared.

We can say, then, that conceptual learning involves focusing on a) aspects of concepts not previously noticed or not previously distinguished from each other or b) conceptual entailments not previously contemplated or not previously thought about clearly and distinctly. It seems quite obvious that our own concepts can be complex enough in themselves or have a complex enough web of entailments that we might learn a great deal by this process, despite the conceptual, i.e. analytic, nature of the truths involved.

Williamson's anti-luminosity argument and infallible knowledge

We have described the issue of fuzzy concepts as "trivial." But in *Knowledge and its Limits* Timothy Williamson has argued that fuzziness is not trivial at all.[52] Starting with the phenomenon of fuzzy concepts and the possibility of indistinguishable experiences, he constructs an argument that purports to show that mental states do not give us a "cognitive home" – that we may not be in a position to know even our own mental states.

Williamson's target is what he terms the luminosity thesis:

> (L) For every case α, if in α C obtains, then in α one is in a position to know that C obtains.[53]

Obviously, some conditions are not luminous. If Sam is admired by his colleagues, he may not be in a position to know that he is admired; his colleagues may be reticent to express their admiration too openly. But traditionally mental states have been taken to be different: if one is in a mental state, then (barring inattention)[54] one is in a position to know that one is in that mental state. According to Williamson, on the contrary, luminosity may fail for some mental state C even if one is in a position to consider or wonder whether C obtains.[55] He clearly means his argument to apply even to situations where one is directly attending to one's experience and wondering whether it has a certain intrinsic character.

The crux of Williamson's argument is a reliability constraint on knowledge that resembles Robert Nozick's notion of tracking the truth. Roughly, the requirement is that one neither believe nor even (apparently) be "fairly confident of" a falsehood in any epistemic situation very similar to a situation in which one has knowledge.[56] That is, if one believes that *p* in a given epistemic situation A where *p* is true, but *p* is false in some epistemic situation A* very similar to A, and if one would either believe or be fairly confident of *p* in A*, then one does not have knowledge in situation A. In this case, A is too "close" to a situation in which the belief in question would be false.[57]

Because it involves the notion of closeness, the notion of reliability is vague. Williamson acknowledges this, but he maintains that some such constraint is intuitively required for knowledge.

> The concept *is reliable* need not be precise to be related to the concept *knows*; it need only be vague in ways that correspond to the vagueness in *knows*. No reason has emerged to doubt the intuitive claim that reliability is necessary for knowledge.[58]

In order to satisfy the conditions for knowledge, therefore, one's belief must have a margin of safety. Williamson uses the analogy of being near the edge of a cliff to illustrate his position. Falling off the cliff represents believing falsely; on his view, knowledge requires that one not believe *p* "unsafely," i.e. when one is too close to the metaphoric edge of believing falsely.

None of this should move the internalist who finds the luminosity of mental states much more intuitively compelling than the reliability requirement. In the previous two chapters we have laid out some of the crippling problems with any such requirements. And in fact, we do not find the claim that reliability in this sense is necessary for knowledge remotely intuitive – *prima facie*, pre-theoretically, it seems absurd on its face. Williamson's anti-luminosity argument depends crucially on an unargued intuition that no internalist should grant.[59]

Nevertheless, one might wonder whether some analogue of Williamson's argument might have force against the infallibilism we defend for *a priori* knowledge. A claim of infallibility, after all, is a very strong claim; and even if the anti-luminosity argument is ineffective against a claim to know *fallibly* one's own mental states, there is at least a possibility that it might work against a bolder claim of epistemic privilege. We are taking on a greater burden by insisting on a higher standard both for foundations and for metafoundations. What might a Williamsonian critique have to say about infallible foundations? And can such a critique be pressed without the reliability assumption?

Williamson's entire argument rests on the notion of gradual change. He imagines, paradigmatically, a case in which S begins the day very cold and gradually, as the morning progresses, becomes extremely warm. In this scenario, S begins by being very confident that he is cold, passes through a stage where he wavers on the question of whether he is cold, and gradually becomes very confident that he is not cold. Williamson points out that, with limited powers of distinguishing sensory states, S might be unable to distinguish his states of what we might call "thermal sensation" from one millisecond to the next.[60] He then brings in the reliability requirement and presses the point that S might believe that he was cold when, in a very similar situation, he would continue having a high degree of confidence that he was cold even though he was not cold according to some "strict standard."[61] S in this scenario can violate safety even when he believes truly; and thus, according to Williamson, he does not have knowledge whenever he is "too close" to being not cold.

But even without the reliability requirement, one might try to argue that when a statement could, in some sense, gradually change from being true to being false by infinitesimal degrees, one could not know infallibly that one was in the state one was attributing to oneself. For would it not be possible that one was really not in that state but in a state resembling it so closely that one could not tell the difference? Is it not plausible that the belief "I'm cold" is fallible – even when taken to refer to one's own sensation and not to some physical fact – because one might be unable to distinguish, *infallibly*, situations where one is cold from those where one is not?

Williamson believes that his type of argument applies very widely. He points out that being appeared to as if a purple patch is ahead can gradually change to being appeared to in some very different way. (He does not specify what the new appearance would be – perhaps as if a tiger is directly ahead.) And it is true that one visual image can be very slowly morphed into another. He also claims that one may not be in a position to know that two words have the same meaning for one, since two words can "gradually diverge" in meaning. "Luminous conditions," he concludes, "are curiosities."[62]

If the argument from gradual change succeeded in showing that there can be no infallible knowledge in such cases, it would have at least some implications both for classical foundationalism and for armchair internalism.

It would be impossible for many sensory foundations to play the role assigned to them by classical foundationalism if being appeared to purple-patchly is not a state that one can be sure one is in. In metaepistemology, if one cannot be sure that two words mean the same thing (or different things) for one, then perhaps many analytic truths cannot play the role we assign them in epistemology.

But Williamson is wrong about the scope of his argument. Take the case of two expressions that mean the same thing for a given subject. In many cases it is not even meaningful to speak of the two words or phrases as gradually diverging in meaning. It makes no sense to imagine oneself as meaning something just slightly different by 'bachelor,' perhaps meaning "someone slightly female" or "someone a little bit married." These terms, at any rate, cannot diverge in meaning by infinitesimal degrees.

Many cases of conceptual containment are also exempt. The concept of *modus ponens* and the concept of truth preservation cannot be thought of as diverging gradually. Either *modus ponens* preserves truth or it does not. Semantic shenanigans with non-indicative conditionals, non-alethic valuations, and degree-theoretic semantics are all changes of the subject. Similarly, the concepts of *a priori* grasp and of infallible judgment are not the sorts of things that could gradually diverge. And to skip ahead to Chapter 7, if direct inference confers a particular degree of confidence given a particular sample size and composition, this conceptual truth does not involve comparing two ideas that might gradually diverge. It is, in fact, almost impossible to think of an epistemic principle we would endorse in which the concepts involved can be thought of as gradually diverging.

Something similar is true of propositions involving conceptual exclusion. For many non-trivial conceptual statements involving exclusion or only partial overlap it is not meaningful to think of the concepts as falling on a continuum or quasi-continuum from one another, so that they might gradually converge in meaning. Consider "A mind is not a physical entity" or "reliable belief production is neither necessary nor sufficient for justification." In neither of these cases can we imagine the two concepts compared as gradually coming closer to one another, so that one might be unable to tell the difference between a mental state where the meaning of one term or phrase excludes the other and a mental state very similar to it where the meanings are actually the same.

Cases in which the relevant concepts do not lie on a continuum are, therefore, exempt from Williamson's argument. But justification comes in degrees, and some degrees of justification may be very similar to others. Does this pose a problem? Not for an armchair internalist who accepts a modal connection between epistemic levels. He can pass the question of the exact level of justification to Johnny Wideawake. Such a perfectly rational and introspectively diligent subject would be able to tell the difference between various degrees of rational confidence that his beliefs might have and would not grant them a higher or lower degree of credibility than was

actually conferred by either the evidence or the foundational connection to truth.

The fact that rational credibility comes on a continuum does not, it should be stressed, mean that there is anything fuzzy about the state of justification that accrues to a belief in virtue of its being based on a particular body of evidence. It is a particularly egregious fallacy to move from a truism about the continuum of credibility to the conclusion that there is, in any interesting or invidious sense, a "mere difference in degree" between the epistemic standing of a belief that is well founded in evidence and the standing of a belief that is founded in nothing at all. If S believes that the Red Sox will win the World Series on the basis of something wholly irrelevant (such as the fact there are many penguins in Antarctica) and nothing else, then S has *no* justification for his belief. The fact that there is something like a continuum of numerical credibilities between this abysmal limiting case and a case in which the belief has a high degree of rational credibility is irrelevant. Suppose that S believes that the Red Sox will win the World Series on the basis of extensive empirical evidence about the Red Sox, their opponents, and the game of baseball. This epistemic state does not lie along a continuum from S's state in the former case, the continuum of numerical probabilities notwithstanding. In fact, there is a difference in kind between the former case and a case in which S's belief is supported by *any* positive evidence. When it comes to having evidence, something is not only better than nothing but also different in kind.

Since reliability has nothing to do with the matter, a reliable mechanism of belief production will not alter this epistemic appraisal. Someone who believes that God exists simply in virtue of the operation of a highly reliable non-inferential belief-forming faculty that is operating in the environment for which it is designed has, *ex hypothesi*, no reasons for his belief and hence no justification; someone who believes that God exists on the basis of a meticulous assessment of the historical evidence for the resurrection of Christ is epistemically distinctly better off, and the more so as the scope of his evidence is wider.

There is also an important discontinuity, at least for strong foundationalists, between foundational and non-foundational empirical knowledge. The referential belief "I am experiencing like *this*" admits of no possibility of error; the inferential belief "I was having a cold sensation at 2 p.m." formed at 5 p.m. on the basis of a printout of temperature readings from a machine attached to electrodes on one's body at 2 p.m. is fallible. This is a difference in kind, not just in degree.

In a wide range of cases, then, there is no significant continuity consideration and nothing resembling Williamson's argument applies. But what about states of sensation or concepts that *can* be imagined as gradually diverging or converging? Must we exclude beliefs about these – and, in particular, beliefs that involve comparing them – from both our sensory foundations and our metafoundations?

Let us consider first beliefs that do not involve comparisons. Williamson's own example, in which a subject believes himself to be cold, is a case of non-comparative judgment. When it comes to judgments about one's own experiential states, the anti-luminosity argument does not touch modern versions of foundationalism in which the experiential state that makes a foundational belief true is referentially constitutive of the belief. When the belief is formed demonstratively, it has the form "I am experiencing like *that*," where the demonstrative picks out the state of one's experience. The subject here need not discriminate his thermal feelings from other thermal feelings that are (or are not) very much like them. He refers directly to the way he actually is feeling at that moment. If he is only somewhat chilly, this does not mean that all he can have is some confidence or fallible knowledge that he is cold. Rather, when he refers directly to his state of experience, he has an infallible belief that he is somewhat chilly. Several recent accounts of foundational knowledge have built in this kind of reference or a corresponding notion of acquaintance that makes the experiential state a constituent of the belief.[63] Such a belief is formed only when the condition in question obtains; but as Williamson himself admits, the anti-luminosity argument does not apply when this is the case.[64]

Williamson tries to anticipate this move by giving an example in which, due to his gullibility, he judges himself to be in intense pain although he is not; the moral, he claims, is that such interdependence of one's mental states and one's judgments about them does not make one's judgments about one's present mental states infallible.[65] But this example misses the point. When a belief is formed like this there is, *ex hypothesi*, no pain to be a constituent of the belief. The claim in question is that beliefs formed *via the relevant sort of reference or acquaintance* are infallible; what secures their epistemic status is the mode of their formation and not just their subject matter. Williamson downplays his own concession regarding beliefs formed only when the condition in question obtains; he avers that such beliefs are very rare.[66] But there are countless sensory beliefs that meet this description because they are referentially formed, and for those beliefs it is simply an epistemic irrelevancy to point out that one could be in a very similar state that one would have trouble distinguishing from one's current state.

At this point, however, we can give Williamson's argument a twist that makes it resemble Wilfred Sellars's argument against the given. Famously, Sellars argues that uncategorized sensations are not contentful and therefore could be of no value as foundations; but on the other hand, in applying a category to one's sensations one opens oneself to the possibility of error.[67] A Williamsonian might argue, similarly, that *interesting* foundational beliefs involve, at least implicitly, a contrast class – not simply "I'm experiencing like *that*" but "I'm experiencing like *this* and *not* like *that*." And indeed, we do think that there are foundational and even metafoundational beliefs that involve contrasts and comparisons. If these involve comparing or contrasting states or meanings that could be thought of as lying along a continuum,

what does this mean for their foundational status? For example, suppose we want to say that one can believe foundationally something like "The way that I am experiencing now is not what I now call 'feeling warm'" or "I'm being appeared to purple-patchly and not tigerly" or, to take a purely conceptual belief, "Coldness is not warmness."

Here it is important to return to the nature of internalist infallibilism of the sort we espouse. Whether the proposition believed involves two states, two concepts, or a state and a concept that can be thought of as lying on a continuum from one another, what justifies the believer is not the *distance* between the two things involved. That is to say, if one believes "This experience is not what I now mean by 'feeling warm,'" one's justification, even one's infallible justification, does not hinge on the fact that one is *very* cold or *at least ascertain distance* from being warm. Trivially, if one cannot by direct acquaintance distinguish one's current state from the concept one designates as "feeling warm," then one cannot know infallibly that one is not warm. But this does not mean that the metalevel explanation of one's justification needs to involve saying, "S is at least this cold, and therefore S is justified in believing that he is cold rather than warm." That would be to reintroduce something like Williamson's safety requirement, a requirement motivated only by reliabilism. What justifies the subject in these cases is the direct act of apprehension that the two concepts, states of mind, experiences, etc. are distinct, *not* the fact – nor, certainly, his apprehension of the fact – that they fall at some particular distance from one another along the continuum in question. It is the nature of his direct act of grasping in virtue of which he is infallibly justified, his acquaintance with the distinctness of the two items. And this holds regardless of how "close" they are to each other.

One may, for example, contemplate one's present experience, with which one is directly acquainted, contemplate at the same time the concept one dubs "feeling warm," a concept with which one is also directly acquainted, and see directly the mismatch between them. One may also see directly that the two concepts – dubbed "feeling warm" and "feeling cold" – are distinct. The fact that states of experience that are hard to distinguish from one another can be thought of as lying between "feeling cold" and "feeling warm" is quite irrelevant to the nature of the epistemic act involved when one grasps directly a truth involving either these concepts or experiences that instantiate them.

There is an interesting parallel here to degrees of complexity in necessary truths. If at a given moment S clearly and distinctly perceives a conceptual truth, this is (trivially) made possible "because" the conceptual truth is simple enough for him to grasp in this way. This is true just in the sense that if the truth were too complex for him to grasp in this way, then he could know it only, if at all, in some other way – through step-by-step demonstration, for example, or by taking someone else's word that it is a necessary truth. It does not follow that he is justified *in virtue of* the fact that

the truth in question falls in a "complexity region" according to some metric of the simplicity of conceptual truths. Nor need he have this metric and know "how simple" the truth in question is, or how far it falls from the edge of the complexity region, in order for his act of grasping to stop a metaregress.

Without a "safety" requirement on knowledge, in short, variations on a Williamsonian theme are no longer pertinent. Seen in proper focus, *a priori* knowledge and foundational sensory knowledge are all of a piece. The introduction of continua is a red herring.

Conclusion

It will come as no surprise that we take the foregoing theory of analytic *a priori* knowledge to be applicable to specific epistemic principles. There are, plausibly, a large number of such principles, and explicating the justification of a given belief might require the invocation of several analytically separable principles, particularly in the case of inferred beliefs. A subject may, for example, be justified in believing that the sun will rise tomorrow in virtue, *inter alia*, of the fact that he is directly acquainted with his own sensory states, that he infers the existence of the external world as the best explanation of those foundational sensory beliefs, and that he makes a further, rational, induction to, say, the conclusion that the sun will rise tomorrow from premises describing the external world (e.g. premises about observed sunrises). If inference to the best explanation (perhaps understood in Bayesian terms) is fundamentally different from induction, there are at least two different types of inference here, so a detailed explication of the subject's justification would include at least three different principles – one explaining how foundational beliefs of the type in question are justified and two explicating the degree of rational confidence conveyed by the inference forms from the relevant premises, held with the relevant degree of rational confidence. A subject who reasons deductively may be following several different forms of reasoning (e.g. *modus ponens*, disjunctive syllogism, etc.), and statements of the validity (and hence, rationality) of each of these inference forms can be regarded as separate epistemic principles.

In the two chapters that follow we intend to show our theory of the analytic *a priori* in action for both deduction and induction. In both cases, we shall give examples of transmissive principles that, on our view, are analytic *a priori* and that, if they are understood clearly and distinctly, vindicate specific forms of reasoning.

6 The Problem of Deduction

Tu quoque?

Epistemically circular reasoning, we have argued, is forceless. In particular, when deployed to show that a mode of reasoning is rational, it offers nothing in the way of genuine grounds for belief since the cogency of the form of the reasoning itself is the question at issue. But at the level of the principles of deduction this stricture seems to raise a problem. How can one be justified in taking the principles of deductive reasoning themselves to be truth-preserving? Surely any cogent argument for the validity of deductive reasoning must use deduction. But will not any such argument, however rigorous, be epistemically circular? And if one is not justified by argument, then how? It appears that in criticizing epistemic circularity, we may have opened ourselves to a damaging *tu quoque*.

William Alston articulates the problem in terms of an attempt to prove the reliability of deduction:

> As for deduction, it quickly becomes obvious that anything that would count as showing that deduction is reliable would have to involve deductive inference and so would assume the reliability of deduction. Just try it. For example, for the case of the propositional calculus, we can demonstrate the reliability of any inference form by truth tables. But doing so is itself a case of deduction.[1]

Alston finds this conclusion disquieting, since it can be generalized to all "sources of belief." Our epistemic principles are finite in number; like the inhabitants of a small hamlet, they cannot take in each other's epistemic laundry more than a few times without creating a circle. Faced with this problem we cannot simply suspend belief, for this would mean suspending all beliefs, and that is not only practically unfeasible but also self-defeating. We are, Alston says, "ineluctably engaged in forming beliefs in ways we cannot non-circularly show to be reliable. And that sticks in our craw."[2]

What sticks in Alston's craw is not, to others, particularly disturbing. According to Plantinga, our inability to justify deduction is of a piece with

the rest of our cognitive life and should be no cause for alarm. Of course deduction cannot be "credentialed" by a derivation of its reliability from any independently certifiable source of beliefs; even God could not do better than an epistemically circular justification of deduction.[3] What we must fall back on is the *spontaneity* of our inclination to believe, despite our awareness that this is, from the standpoint of traditional skepticism, a bruised reed:

> When we contemplate the corresponding conditional of *modus ponens*, we just find ourselves with this powerful inclination to believe that this proposition is true, and indeed couldn't be false. But (as we also know) such inclinations are by no means infallible. We really don't have any reasons or grounds for this belief; we simply, so to say, *start* with it.[4]

Some philosophers find epistemic circularity at the level of allegedly *a priori* knowledge positively welcome. Susan Haack concludes a detailed critique of deductive justifications of deduction with a distinct air of satisfaction:

> The moral of this paper might be put, pessimistically, as that deduction is no less in need of justification than induction; or, optimistically, as that induction is in no more need of justification than deduction. . . . Those of us who are sceptical about the analytic/synthetic distinction will, no doubt, find these consequences less unpalatable than will those who accept it. And those of us who take a tolerant attitude to nonstandard logics – who regard logic as a theory, revisable, like other theories, in the light of experience – may even find these consequences welcome.[5]

The easiest way to see the difficulty posed when one attempts to justify the rules of deductive inference is to look at the structure of a plausible justificatory argument. Taking a cue from Alston, let us try to construct an argument from the truth table for conditional propositions to the inevitably truth-preserving nature of *modus ponens*. First we display the relevant information in a truth table:

P	Q	(P→Q)	((P & (P→Q))→Q)
T	T	T	T
T	F	F	T
F	T	T	T
F	F	T	T

Now we are in a position to argue as follows:

(A) 1. Every line of the truth table that assigns truth to both P and (P→Q) also assigns truth to Q.
2. But if this is so, then whenever P and (P → Q) are true, Q must also be true.

Hence,

3. Whenever P and (P → Q) are true, Q must also be true.

Here, line 3 ascribes to *modus ponens* the property that is the very definition of deductive validity, so we seem to have achieved our goal. Alas! (A) itself is an instance of *modus ponens*, and the argument therefore runs afoul of our ban on epistemic circularity. The extension of this problem to roundabout deductive justifications of deduction is straightforward. For every rule of inference used in the justification, a thoroughgoing deductive skeptic will demand an independent justification. If we are resourceful enough never to repeat an argumentative pattern, we can prolong the discussion but never resolve the issue to the skeptic's satisfaction; if we are not so resourceful, we will eventually repeat a form, creating an epistemic circle.

To the student of logic, this raises some disquieting questions about the fundamental results of logical metatheory. One of the showpieces of a rigorous course in elementary logic is the demonstration that both the propositional calculus and standard first-order predicate calculus are *sound* and *complete*. Put briefly and somewhat colloquially, a proof of *soundness* assures us that our rules of inference are not too strong, and a proof of *completeness* assures us that they are not too weak. In view of the inevitable use of inferences in these metatheoretical pursuits, as Michael Friedman notes, this procedure is "in an important sense circular." Nevertheless, he goes on, it

> does nonetheless provide an important kind of justification for deductive methods of inference. It shows that there is a desirable harmony between the methods we use in practice and our conception of what the point of those methods is, namely, preservation of truth from premises to conclusion. ... The completeness theorem, therefore, accomplishes something significant despite its circularity.[6]

But suppose that unbeknownst to us our rules of inference *are* too strong, allowing us to derive conclusions that we would have been unwilling to sanction if we had been sufficiently vigilant. Then following those very patterns of inference we might in all innocence derive things that did not, in point of logical fact, really follow from our premises; and why should we not find, among the results that we might thus derive, the soundness and completeness of our system of logic?

Consider the following argument for the truth-preserving nature of the well-known deductive fallacy of affirming the consequent, known by weary logic professors as *modus morons*:

(M) 1. If the argument from (P → Q) and Q to the conclusion P is valid, then (P v ∼P) is a tautology.
2. (P v ∼P) is a tautology.

Hence,

3. The argument from (P → Q) and Q to the conclusion P is valid.

The premises of (M) are true; and though its form is an instance of *modus morons* itself, Friedman's line of reasoning would encourage us to let that pass. Yet it ought to go without saying that this argument does not accomplish anything significant and certainly does not constitute a justification for affirming the consequent. If this argument gives us a feeling of "desirable harmony" between the practice of affirming the consequent and the truth-preserving nature of deductive reasoning, so much the worse for us.

The moral to be drawn here is not that logical metatheory has no point or that soundness and completeness are not, after all, desirable properties. But Friedman is simply mistaken to think that such arguments can provide, as he claims they can, "an important kind of justification for deductive methods of inference." Rather, the moral is a conditional one: if we cannot trust the specific forms of reasoning used in our metatheoretic arguments, then our derivation of soundness or completeness is pointless. Far from giving us a justification for deductive inference, such a derivation is an exercise in blind faith.

Intuition, demonstration, and the status of metatheory

In the previous chapter we expounded and adapted a Lockean theory of intuition to analytic knowledge generally. Logical metatheory requires yet a further Lockean concept, the notion of demonstration. When we turn to the relation of intuition and demonstration, we find Locke adamantly insisting on the primacy of the former. Both in the *Essay* and in the controversy with Stillingfleet he maintains that the syllogism, though not to be despised, is itself dependent on intuition – that the latter is, in Aaron's felicitous phrase, the "cognitive core of reasoning as inferring."[7] A famous passage added to the fourth edition of the *Essay* spells this out in detail.

> But God has not been so sparing to men as to make them barely two-legged creatures, and left it to Aristotle to make them rational, i.e., those few of them that he could get so to examine the grounds of syllogisms, as to see that, in above three score ways that three propositions may be laid together, there are but about fourteen wherein one

> may be sure that the conclusion is right; and upon what grounds it is, that, in these few, the conclusion is certain, and in the other not. God has been more bountiful to mankind than so. He has given them a mind that can reason, without being instructed in methods of syllogizing: the understanding is not taught to reason by these rules; it has a native faculty to perceive the coherence or incoherence of its ideas, and can range them right, without any such perplexing repetitions.[8]

Despite some strong language, Locke does not disparage the syllogism *tout court*. In his polemical writings against John Edwards just two years before the insertion of this passage in the fourth edition of the *Essay*, Locke calls syllogism "the true touchstone of right argument" and intimates that Edwards's work fares poorly when measured against such a standard.[9] The point rather is that the study of syllogism cannot supply a grasp of conceptual relations if such grasp is lacking in the first place, and that often the purposes to which syllogistic reasoning is put, whether persuasive, pedagogic, or dialectical, are better served by indicating the containments and connections of concepts directly.

In Chapter 5 we discussed various intuitions of implication as reasonable extensions of the Lockean notion of conceptual containment; these intuitions form the basis for the construction of demonstrations. In demonstration, formal proofs are built by the application of rules that license the transition from a set of formulas to a formula. The application of a Lockean epistemology to formal logic will involve the certification of some such rules in virtue of the meanings of the component terms.

The most natural way to display those meanings is to use truth tables. But as we saw above, to *argue* from the truth table of the corresponding conditional for *modus ponens* to the truth-preserving nature of this inference by means of the schema (A) would entangle us in epistemic circularity. Yet every teacher of logic knows that some truth-table arguments are both possible and pedagogically helpful. How can we reconcile epistemic rigor with pedagogic practice?

The key term here is 'display.' Truth tables are an aid for the logically myopic, not a cure for the logically blind. We display the meanings of connectives in a matrix in order to clarify their use – to stress, for example, that the formal symbol '&' is not being used in a temporally sensitive way as its English counterpart 'and' sometimes is, that the symbol for 'or' is being used inclusively, that negation cannot here be doubled as an intensifier. There is nothing epistemically circular about this procedure. But any arguments we make *from* truth tables to the virtues of rules of inference presuppose that the reader already grasps the entailments in question; and when those entailments turn on the logical properties of terms like 'and,' 'or,' and 'not,' we are presupposing that the reader has a sufficient grasp of those concepts to see the entailment.

As myopia comes in degrees, so, as we have repeatedly emphasized, does logical perspicacity. There may be the occasional genius who can see at a glance that a complex formal system is sound or complete, but for most of

us this is a bit of knowledge arrived at by demonstration and not something apprehended in a direct intuition. And this is precisely why we *need* soundness and completeness proofs. Step by intuitively certified step we can lay out the argument that our rules of inference are not too strong or too weak. Demonstrations, then, are in effect *explications* which draw out at least some conceptual entailments of premises for those who can grasp clearly and distinctly those premises by themselves.

Locke held that demonstration led to certainty just as did intuition,[10] but this claim does require a caveat. Whether or not demonstration brings rational certainty will depend in any given case upon whether memory is involved. In some cases the very act of demonstration may show conceptual entailments so clearly that, once the subject sees the demonstration, he is then able to hold the entire chain of reasoning in his mind at once without reliance upon memory. Such a mental act will be closely akin to intuition, although it will be, so to speak, complex rather than simple, in that it will consist of the immediate perception of a set of logical relations that reveal conceptual connections rather than of a single conceptual relationship. On the other hand, if the demonstration can be seen only partially at any one time, the subject will be relying on memory and so the resulting conclusion will not be entirely a matter of *a priori* knowledge. However, as we discussed in both Chapter 2 and Chapter 5, this fact does not impugn the infallible nature of *a priori* knowledge as such.

What difference does the Lockean analysis make to the epistemic question raised about the status and putative epistemic circularity of logical metatheory? The rationality of the modes of reasoning used in an epistemically cogent consistency or completeness proof must themselves be intuitable – fixed points on which the proof can turn. The status of those fixed points is not in question here, any more than it is in question when we use them to explain how the truth-tables convey the meaning of the logical connectives. Hence, there is no metaregress and no epistemic circularity, for intuitable truths satisfy the requirements of the modal principle discussed in Chapter 4.

For example, in Nagel and Newman's charming proof of absolute consistency for an axiomatized system S of sentential logic,[11] we find the following argument:

(P) 1. If there is a formula expressible but not derivable within system S, then system S is consistent.
2. There is a formula expressible but not derivable within system S.

---

3. System S is consistent.

This argument would be epistemically circular if it were offered as a reason for the validity of *modus ponens*; for (P) is itself an instance of *modus ponens*, which is a rule of the system under consideration. But this is not the point of (P). Rather, the metalogician offers (P) on the understanding that the reader can see, intuitively, the legitimacy of *modus ponens*.

That is not to say that the metatheoretic arguments lack cogency. Rather, they are cogent for those who are able to understand them as they are intended. Nagel and Newman's proof *establishes* the absolute consistency of S. But the consistency of S is not something any ordinary person would plausibly find obvious, and few even among logicians could hold it in the mind in a single act of intuition. This is why, in the fuller version of the proof, the various steps undergirding the premises are chained together in a demonstration: the verification that each axiom is tautologous, the proof by mathematical induction that each rule of inference transforms tautologies into tautologies, the demonstration that from two contradictory formulas every formula is derivable, the exhibition of a statement that is not a tautology. The proof proceeds by just those steps that we require with our limited logical perspicacity. And as we understand them and the way that they fit together, we have demonstrative knowledge in virtue of our ability to grasp the cogency of the fundamental steps.

Are there people so inept logically that they cannot intuit the validity of *modus ponens?* In an academic world where postmodernists are viewed as intellectual superstars, it would be rash to be sure that there are not. But it would be equally foolish to suggest that they could profit from a consistency proof, or that the genuine cogency of *modus ponens* is in any way called into question by the existence of such cognitive cripples. Perhaps persistent work on logic could act as therapy to improve their cognition. But unless and until they develop the ability to recognize the validity of basic rules of inference, logical metatheory must remain for them a closed subject.[12]

The notion of logical intuition is what is missing from Friedman's account of metatheory, and the lack of it is what makes his account dissatisfying. Only with a concept of *a priori* intuition do we have the resources to say why an alleged proof of anything using *modus morons* is worthless, since the validity of *modus morons* is not intuitable and, indeed, the fact that it is *not* valid *is* intuitable, if only by clearly grasping the relevance of a counterexample. And only with this account do we have the resources to explain how the *a priori* answers the problem of the metaregress for deductive logic. The fact that *modus ponens* is a valid form of inference is itself an intuitable analytic truth, as are similar principles for other deductive argument forms. Such intuitions will involve understanding concepts like validity, truth, and material implication.

The fact that we can demonstrate the consistency or completeness of deductive systems by metatheoretic proofs does not mean that such proofs are needed to *defend* basic logical steps like *modus ponens*, as if the rationality of these steps were intrinsically dubitable and stood in need of an argument to show their validity. Rather, metatheory demonstrates the properties of entire systems by way of steps whose validity we are able to grasp directly. Intuitable logical truths thus provide the metafoundational stopping places for deduction.

Objection: Fallible logical knowledge?

On Susan Haack's view, fallibility is in the first instance a feature not of propositions but of persons; for she acknowledges that if there are any genuinely necessary principles of logic then it is not possible that *they* are false, whatever we may think of them. Logic may be unrevisable because necessarily true, but what we *take* to be the necessary truths of logic is, according to Haack, another matter entirely. But an infallibilist position does not depend upon a widespread human capacity for infallible logical knowledge. Instead, we are claiming that all genuine *a priori* knowledge, including knowledge of logical truths without any admixture of memory, induction, and the like, is by its nature infallible. Haack's position seems to be that all our logical beliefs should be taken as in principle revisable because we cannot tell that any of them correspond to the "real" logical truths. In other words, her position is that *all* logical knowledge is fallible.[13]

Haack's arguments for this general fallibility are various, and many of them are of the sort that we have dealt with in Chapter 5 – arguments from human errors about necessary truths, arguments against self-evidence as an indicator of truth, and so forth. But Haack contends in addition that the multiplicity of logical systems supports logical fallibilism:

> Another reason against epistemological over-confidence is the knowledge that other people hold, with as much confidence, beliefs incompatible with one's own. And this motive operates in the sphere of logic, too; the very plurality of logical systems speaks against our possession of any infallible capacity to ascertain the truths of logic.[14]

Here, Haack *assumes* the falsity of a traditional position concerning the infallibility of *a priori* knowledge. She issues a call to humility by way of the fact that others disagree with us. But if a subject has the relevant conceptual grasp, he will not *and should not* be rationally moved by such considerations. Strictly speaking, the question at issue is whether a subject who possesses a clear and distinct *a priori* grasp of a logical truth is thereby justified in an infallible way. Anyone who is at a given moment grasping a truth in this fashion need not be troubled by the historical fact that people have sometimes been wrong about logic.

In fact, the existence of a "plurality of logical systems" need not even mean that some of the people advancing such varying systems must be wrong about logic. Suppose that Jack and Jill disagree about whether the sentence "all cats are animals" is true. Must we attribute to either of them a fallible capacity to ascertain analytic truths? This is not forced on us by their disagreement, and the Lockean view provides us with the resources to see why. Failure of logical insight is a possible explanation for their disagreement, but it is more plausible to suppose that Jack and Jill are working with alternative meanings of 'cat' or 'animal.' Their disagreement is the best evidence we could wish for on this point.

The same goes, *mutatis mutandis*, for a dispute regarding a purely logical statement such as the law of excluded middle:

$$\vdash (P \vee \sim P)$$

Even though no natural language term like 'cat' is included here, there is still room for difference in interpretation of the logical symbols typically construed as indicating disjunction and negation, and there is also room for various strictures on the sorts of things 'P' may stand for in such a schema. Despite all of the ink spilled over the law of excluded middle, there has never been a plausible challenge mounted against the logical necessity of '(P v ∼P)' that has not involved an alteration in the meanings of one of those connectives, the domain over which 'P' ranges, or the meanings of the valuations we ascribe to the formulas of the language. Instead we find proposals for alternate truth tables or axiomatic replacements for truth tables where, as in Heyting's intuitionist logic, the connectives are not interdefinable and there is no finite characteristic matrix.[15] But what could make it clearer that what is being proposed is a change of the meanings of terms?

Yet even here we must tread cautiously. For as Haack has persuasively argued, a demonstrable change in the meanings of the connectives is not by itself sufficient to guarantee that two logical systems are not rivals; for they might be intertranslatable, and in that translation we might discover that the deviant logician still denies sentences that are theorems for the classical logician.[16] What is clear, however, is that mere disagreement over notationally identical formulas is insufficient to guarantee rivalry when the connectives have changed in meaning. Those who want to put forward a logical system that is to be a genuine rival to classical logic need to do more than change the classical truth tables.

How much more? Haack finds this question difficult to answer, and it is hard to pin down her notion of "real" rivalry. She suggests that "straightforward rivalry" may be conceived as deviance "unaccompanied by any meaning variance,"[17] but the scope of the meaning variance in question is unclear. Should van Fraassen's supervaluational languages, in which all (classical) tautologies are assigned the value T, all (classical) contradictions F, and all others are left without valuation, count as deviant? Haack thinks this is possible, but she also concedes that such systems may be regarded as "semantically non-standard"[18] and later characterizes them without qualification as "semantically deviant."[19] Meaning, apparently, has varied.

It is in one sense easy enough to offer a sufficient condition for genuine rivalry: a deviant logic DL is a real rival to classical logic CL if, in DL, one of the rules of inference of CL, *understood as it is intended in CL*, is invalid, *as invalidity is understood in CL*; or that a theorem of CL, *understood as it is intended in CL*, is not a theorem, *as theoremhood is understood in CL*. Someone could *say* that *modus ponens* is invalid, just as someone can *say* that excluded

middle is false as classically understood. Here would be deviance of the most dramatic sort – postmodern deviance, so to speak.[20]

But Quine's principle of charity suggests that we ought not take the postmodern logician seriously. What better reason could we have to believe that the would-be radical is merely confused, at least confused about classical logic if not mentally deranged at some deeper level? Here logical monism, which Haack defines as the position that "there is just one correct system of logic," is correct, provided that it is properly qualified. With respect to the domain of propositions and the classical truth-values, given the meanings classically ascribed to the connectives and the classical concept of validity, classical bivalent logic is indeed the only correct logical system. Attempts to deviate from it without changing the meaning of one or more semantic components are serious only in the sense that they darken counsel and may confuse the unwary; they are not to be taken seriously.

As Haack herself acknowledges, this is not our situation with the deviant logics on offer. A relevance logician, for example, will deny the validity of *modus ponens* even when the conditional is understood in the material sense. But as Haack points out,

> what he means, when he says that MPP *isn't* valid, isn't what the classical logician means, when he says that MPP is valid, since the relevance logician would agree that MPP is valid in the classical sense of '*valid*'.[21]

This is not deviance unaccompanied by any meaning variance; it is merely a shifting of the meaning variance from the object level to the metalevel. The relevance logician means something else by 'validity.' Well and good; we can explore other metaconcepts. But where has the rivalry gone?

It may be urged that deviance at the level of metaconcepts is precisely what constitutes real rivalry. Haack suggests that in the case of relevance logics we are dealing with diverse attempts to capture, in a formal way, our extra-systematic idea of validity. But from a classical point of view it seems clear that there is not just one (somewhat fuzzy) notion being explored by different logicians in different ways; there are two notions, clearly distinct and serving different purposes. If you ask a classical logician what can be validly derived from inconsistent premises in classical bivalent logic, he will tell you that you can derive any formula whatsoever; but if you ask him what someone with inconsistent premises should believe, he is likely to tell you that the inconsistent subject should, in Ogden and Richards's famous phrase, amend the dilemma. There is no double standard here; the classical concept of validity is not intended as a doxastic mandate for the discoherent. If this sort of doxastic mandate is what the deviant logician seeks, then he has changed the subject.[22]

A suspicion may remain that this brisk treatment of deviant logics does not do justice to the earnestness with which the alternative logics are

advanced. If they are not rivals to classical logic, then why do so many people urge that they are and that we must learn to think in new ways? Haack suggests that deviant logics may, in some cases, be construed as "alternative formal projections of the same informal discourse."[23] It is famously debatable whether and to what extent material implication captures the notion expressed by the "if ... then" construction in English. Might there not be a substantive issue regarding the proper representation of natural language arguments?

To be sure, we can devise a logic that captures the temporal connotations of 'and' or the plurality implied by 'some.' Here, however, Haack's own analysis tells against what we might call *macho* pluralism – the position that there are genuine rivals to CL that deserve serious consideration. For we may view formal representation of various aspects of informal discourse, in Haack's own words, as abstracting from what we take to be "irrelevant or unimportant features of informal discourse."[24] But in that case we are not representing the informal argument whole; we are representing a collection of its features. If someone impressed with other features wishes to construct a system in which they can be represented, then by all means let him do so. But he will be representing a different subset of the characteristics of the informal argument, not representing the same characteristics in an incompatible way.

Classical bivalent logic is a many-splendored thing. It comprises concepts of truth and falsity, validity and consequence, valuation and various connectives. We may, sometimes with profit, explore formal systems that invoke non-classical concepts as replacements for any of these. Such systems need not be *wrong*. There are all manner of interesting mathematical and logical structures that can be defined by internally consistent constraints on the interpretation of their symbols. A traditional position has no difficulty accommodating this fact; it is consistent with what we might call modest logical pluralism, the position that differing meanings for the logical connectives, differing constraints on the domain of the propositional letters, and so forth may yield differing structures of legitimate intrinsic interest. On the Lockean view this amounts to adopting a new notion of, say, disjunction, or admitting a new range of valuations, or restricting or enlarging the set of items for which the sentential variables may stand; and it is no more surprising that one should get different-looking theorems from such a change of interpretations than that one should assent to different sentences in view of a decision to denote some new and different concept by the term 'bachelor.'

One restriction, however, applies; to be intelligible, such systems must be developed *consistently*. We may express this in various ways, by speaking of the constraints imposed by the new interpretations, for example. But whatever our verbal formulations it seems that there is no escape from a metalogic that is, broadly speaking, classical and bivalent. Donald Williams puts this trenchantly:

> [A]lthough we can invent artificial systems of logistic *ad infinitum*, whatever postulates, definitions, formation rules and transformation rules we lay down, with whatever language or metalanguage we operate, their consequences are what they are, irrespective of convention and convenience, by the one inescapable sort of necessity.[25]

Conclusion

Where does all of this leave us with respect to the problem of deduction? When one is actually grasping the corresponding conditional of *modus ponens* or the law of excluded middle, as understood classically, one can see clearly and distinctly, by Lockean intuition, that these are and must be true and that one is justified in believing them. In those instances, the existence of (supposedly) alternative logical systems is irrelevant to the proposition held before the mind. If the subject, grasping the law of excluded middle, is told that some alternative system denies that law as it is classically understood, he will immediately see that, at least in that respect, such a system must be incorrect. The individual propositions of classical bivalent logic, held in the mind and understood clearly and distinctly, stop the metaregress for deduction.

But there may be times when we are not contemplating those propositions in and of themselves, or when we are seeing only some of them clearly, and when we have questions about the system CL *as a whole*. In that case, proofs that proceed by steps we *can* grasp *a priori* can show the completeness and consistency of the system by demonstration. And if we can hold a demonstration clearly in our minds all at once without relying on memory, we can know its conclusion by a complex intuitive act.

Furthermore, if we are trying to explain the phenomenon of logics that purport to be serious rivals to classical logic, we can do so by seeing that such logics (if they have any value whatsoever) involve changes in meaning for the propositions allegedly denied in the alternative systems. This process may be helpful to those who have *never* grasped the relevant propositions and who are wavering or being led into confusion by the purported existence of rival systems. And it is helpful as well for anyone who is not at a given moment focusing on a specific proposition but is discussing in the abstract the question of whether classical logic could be incorrect or revisable. Thus by leading us back to concepts and meanings, the defense of classical logic against claims of rivalry reinforces our understanding of the central importance of analytic *a priori* intuition.

7 The Ground of Induction

Any traditional, internalist epistemology must come to terms with the question of non-deductive inference. In our case, the problem of defending non-deductive inference is particularly urgent since we are debarred by our own scruples from taking some of the more popular positions on this topic. We cannot appeal to an inductive justification of induction without running afoul of epistemic circularity; we cannot reduce the problem of induction to the meliorative project of giving epistemic advice without erasing the distinction between success and rationality, losing what is distinctive to internalism. If we are to have any hope of retaining induction as justificatory, there is no alternative for us but to tackle Hume head-on and argue that induction, which C. D. Broad memorably described as "the glory of science" and "the scandal of philosophy,"[1] is defensible *a priori*.

This is a decidedly iconoclastic position. It would be fair to say that the consensus among contemporary philosophers is that Hume's problem, or at any rate a refurbished modern version thereof, is quite simply and clearly insoluble.[2] We will argue that, to the contrary, direct inferences are underwritten by the metafoundational insight of the proportional syllogism, and that individual direct inferences can be justifiably based on Bernoulli's theorem. These tools provide the key to an internalist refutation of Humean skepticism regarding induction.

Hume and "Hume's problem"

Hume's problem of induction is, *prima facie*, quite straightforward and constantly repeated: since the falsehood of any non-demonstrative proposition is clearly conceivable, how are we, on the basis of our experience, to infer the truth of any "matter of fact" that lies outside our experience? Lacking demonstrative proof, we find our reasoning attended with some uncertainty; and, says Hume, if we argue by factual reasoning that we can trust our past experience, we will be begging the question by assuming that nature will remain uniform.[3] Though we customarily do draw conclusions regarding unknown matters of fact, Hume offers us in the end only a

psychological explanation for this practice rather than the rational justification for which we might have hoped.

A salient passage in the *Abstract* exhibits Hume's critique of inductive inference succinctly:

> 'Tis evident, that Adam with all his science would never have been able to demonstrate, that the course of nature must continue uniformly the same, and that the future must be conformable to the past. What is possible can never be demonstrated to be false; and 'tis possible the course of nature may change, since we can conceive such a change. Nay, I will go farther, and assert, that he could not so much as prove by any probable arguments, that the future must be conformable to the past.[4]

Hume's strongest claim in this passage, which constitutes the core of Humean challenges to induction, is that no "probable arguments" will underwrite the conclusion that the future resembles the past, and hence, according to Hume, no such argument can underwrite induction. But beneath the apparent simplicity of this challenge lie several presuppositions that must be unearthed and reconstructed before the basic problem can be seen clearly.

Despite its modern sound, Hume's phrase "probable arguments" does not refer to arguments in which the premises provide less than a guarantee for the conclusion; rather, Hume uses the phrase to denote *deductive* arguments with contingent premises.[5] Recognition of this fact clears up several puzzling features of Hume's discussion. The emphasis he lays on the Uniformity of Nature, for example, makes sense from this perspective: it is an attempt, ultimately unsuccessful, to insert a contingent lemma between the premises and the conclusion in order to create a valid deductive argument. Similarly, his reiterated claim that a connection between matters that lie within our experience and those that do not cannot be found *a priori* becomes clear and even obvious on a deductive model: adding a necessary truth to the premises will not strengthen the argument.

The interpretive point that Hume is working with a limited conception of "probable inference," though hardly original, has important consequences for the problem of induction. It has become obvious in the 250 years since the publication of the *Enquiry* that this restriction, taken literally, would render any reasonable solution to the problem of induction impossible by ruling out the possibility of learning from experience. Non-demonstrative inference, even granting *arguendo* that it is epistemically legitimate, is not in general monotonic: adding further premises to a non-demonstrative argument may drastically undermine support for the conclusion. Non-monotonicity is a feature we need for a plausible theory of inductive inference. Precisely because (as Hume never tires of pointing out) our conclusions may be false despite the favorable data, we must be able to weaken the

support for those conclusions by enlarging our experience – discovering, say, a black swan or a white raven. Deductive inference, however, is monotonic: adding further premises cannot invalidate a deductive argument. Any interesting version of the problem of induction must leave open at least initially the possibility of genuinely justificatory non-deductive forms of inference.

Such a reconstruction of Hume's challenge casts the problem of induction in a different light. For one thing, it renders pointless Hume's own dilemma for inductive inference – that the premise added to the argument would be question-begging if it stated a matter of fact but useless if it expressed a relation of ideas. Once the skeptical problem is reformulated to allow in principle for a non-deductive inference form, no such dilemma arises. In particular, Hume's rationale for insisting that one must *assume* the uniformity of nature to argue that the sun will rise tomorrow has now disappeared.

But something of the spirit of Hume's dilemma survives at the metalevel. The *epistemic principle* that connects the premises and conclusion of an inductive argument must be in some sense rationally defensible: but how can any such defense be mounted? In particular, if the defense of the epistemic principle itself depends essentially on contingent features of our world lying beyond our immediate experience, then the problem of induction engenders a metaregress: for by what means are we to infer that our world possesses the contingent feature in question? If, on the other hand, the epistemic principle is to be defended wholly *a priori*, then it is incumbent on the epistemologist to produce a purely logical or mathematical principle adequate to the task. And in view of the vagaries of the physical universe, which owes us no cooperation in our inductive enterprise, this has seemed to many inductive skeptics about as promising as an attempt to square the circle.

One way of putting the skeptical challenge regarding any form of inference is the question: "Granted that these premises are true, and that this inference form links them to the indicated conclusion; granted also that I prefer truth to falsehood; why should these facts commend the conclusion to me?" For deductive argumentation, we appear to have a simple and gratifying answer: "Because it is guaranteed that in accepting the conclusion you will *never* believe falsely." And, as we have argued in the previous chapter, the truth-preserving nature of deductive principles is certifiable *a priori*. But precisely because the premises of a non-deductive inference form do not guarantee the truth of the conclusion, it is difficult to justify what *appears* to be the parallel response: "Because it is guaranteed that, in accepting the conclusion, you will *usually* believe truly."[6]

Yet we should be suspicious of the demand for guarantees of success, even of success most of the time. In Chapter 2 we distinguished the notion of an "intrinsic" connection to truth from a reliabilist or extrinsic connection and argued that non-deductive arguments purport to provide an intrinsic connection to truth by way of the concept of rationality. To insist on a guarantee that we "usually believe truly" when following non-deductive

arguments is to require an extrinsic connection; and that is a requirement we are entitled to reject. As we will see shortly, the distinction between success and rationality, extrinsic and intrinsic connections to truth, lies at the heart of the problem of induction. What is required to address the legitimate question in the skeptical challenge is a metaprinciple of non-deductive inference that provides an intrinsic connection to truth.

Hume's challenge is as interesting for what he allows as for what he contests. He takes it for granted that in situations where we wish to engage in inductive extrapolation we may unambiguously identify the right reference class – sunrises and loaves of bread, for instance. In this he shows a surprising streak of common sense. For although the selection of a reference class is not always a straightforward matter and sometimes creates serious problems, there do seem to be many cases in ordinary life where a single reference class strikes us as obviously right and all rivals as obviously wrong. Justifying our spontaneous preference in such simple cases is itself a significant philosophical undertaking. But it is not Hume's problem.[7]

Many of Hume's examples of inductive extrapolation have an additional feature that seems to make them particularly difficult candidates for justification: they move from premises in the past and present tense to a conclusion regarding the future. Opponents of direct inference have seized upon this point to create a modal barrier against any statistical extrapolation into the future. If it is ever legitimate to argue from the properties of a sample to those of an unsampled member of the population, so runs the objection, this is only because each member of the population had an equal chance of being selected in the sample. But by definition, one *cannot* sample the future. Future ravens had no chance to be included in one's evidence; the probability that a raven not yet conceived is included in any past sample is simply zero. The sample is not, as a result, "truly random" with respect to the set of all ravens, and it therefore has no bearing on the future.[8] Or so runs the argument.

We have, then, what we may call a general and a specific problem of induction. The general challenge is to produce a method of inference in which, although the premises do not logically entail the conclusion, they do render it genuinely probable in an epistemically significant sense. The specific challenge grants that the problem of the reference class is often unproblematic but requires a method of inference that sanctions extrapolation from a uniform sample to the presence of the property in question in an unexamined member of the population. This problem has, as a particularly interesting special case, the projection of past data into the future.

Direct inference and the problem of induction

A long tradition, stretching from Bernoulli and Bayes to Howson and Urbach, identifies the inference from sample to population as an exercise in "inverse" reasoning, of which Bayesian reasoning is the modern form.

Unfortunately, as early critics of inverse inference were quick to point out, it is possible to approach induction in this fashion only if one has a prior probability regarding the proportion of, say, black ravens among ravens generally. How such priors are to be acquired is the fundamental problem of Bayesian inference; its apparent intractability is doubtless the chief stone of stumbling for non-Bayesians.

We do not propose here to survey Bayesian responses to this difficulty, much less to adjudicate disputes about their adequacy. What we want to investigate instead are the prospects for a very different approach to inductive extrapolation that does not invoke prior probabilities and inverse inference and that therefore has some claim to be the more fundamental form of reasoning. This approach utilizes the appearance of a feature (e.g. being black) within a sample of a population (such as ravens), *direct* inference, and Bernoulli's theorem to calculate the probability that the feature appears with a particular frequency within the population as a whole. The method, if defensible, should hold interest for anyone concerned with the problem of induction.

Direct inference is perhaps the simplest and most natural expression of a "degree of entailment" interpretation of probability. Given that the frequency of property X in a population G is p, and given that a is a random member of G with respect to possession of X, the probability that a is an X is p. We will return to the requirement of randomness in due course.

The intuitive appeal of direct inference comes out strongly in simple examples. Donald Williams, a passionate advocate of direct inference, describes it in terms of the "intermediate cogency of proportional syllogisms."[9] Just as the classical syllogism warrants our concluding, from

1 All G are X
2 a is a G

with full assurance, that

3 a is an X,

and from

1 No G is an X
2 a is a G

with full assurance, that

3 a is not an X

so the proportional syllogism, subject to the restrictions mentioned above, licenses our inference from

1′ m/n G are X
2′ *a* is a G

with assurance m/n, that

3′ *a* is an X.

The first two syllogisms given here are of the classical forms Barbara and Celarent. But as Williams points out, we use the classical syllogism but rarely: our major premises are not of the form "All falling barometers portend storms" or "All red-meated watermelons are sweet" but rather the more modest form that falling barometers *generally* portend storms and *most* red-meated watermelons are sweet.

> In the cadres of the traditional deductive logic, these changes make a fatal difference: the propositions that falling barometers generally portend a storm and that the barometer is now falling entail, strangely enough, nothing whatever about an impending storm. ... Impatient with this finicking of the logician, the native wit of mankind has jauntily transcended the textbook formulas, has found the principle self-evident that if *All M is P* makes it certain that any one *M* will be *P*, then *Nearly all M is P* makes it nearly certain, and has quite satisfactorily predicted its storms and purchased its melons accordingly.[10]

The notion of the statistical syllogism as a generalized form of the traditional one admitting intermediate grades of logical cogency is attractive, and a substantial number of philosophers have incorporated something like it in their treatment (though not always their justification) of inductive inference.[11] Indeed, Barbara and Celarent can readily be seen as limiting cases of the proportional syllogism when m = n and m = 0, respectively.[12] From this point of view, statistical syllogisms constitute a spectrum of inferences, each moving from statistical information to singular statements about members of the relevant class. The conclusion, as in the traditional syllogism, is always categorical, but the level of confidence varies with the proportion cited in the major premise.

It is the proportional syllogism that provides a transmissive epistemic principle capable of stopping the metaregress for non-deductive inference. It is no accident that it spans the logical space between the two traditional syllogisms, nor does the randomness requirement vitiate the comparison. In Barbara and Celarent, symmetry is guaranteed because the major premise puts all of the individuals in question on an even footing: in Barbara they all have the property in question, and in Celarent none of them has it. In the proportional syllogism, epistemic symmetry needs to be invoked as a metalevel constraint to prevent absurd misapplications; when some members of a population have, and others lack, a given property, it is obviously

unreasonable to invoke mere proportions if we have concrete information about the individual member in question. But under conditions of epistemic randomness that we will examine in detail below, a direct inference that proceeds according to the proportional syllogism is as evidently rational as any standard deductive inference.

Granting for the moment that the randomness constraints on such an inference are satisfied, we have still to account for our knowledge of the major premise. How can we come by the knowledge that m/n ravens are black? In particular, how are we to come by it in a fashion that does not examine all ravens *seriatem*, including the one named *a*, so that in the last analysis direct inference falls prey to an analogue of Sextus Empiricus's complaint about the traditional syllogism – that to complete the enumeration required to establish the major premise, we will have to make use of the conclusion, thus rendering the subsequent argument circular?[13]

It is true that we cannot *guarantee* the major premise without examining all of its instances. But as Williams points out, we can obtain a rational credibility for the major premise by a clever combination of Bernoulli's "law of large numbers" and a second direct inference. Crudely but briefly put, Bernoulli's theorem says that most logically possible large samples differ but little from the population out of which they are drawn – where "most" indicates a satisfyingly high percentage and "little" a gratifyingly small deviation from the true value, provided that "large" is sufficiently great.[14]

Armed with an n-fold sample of balls (from the statistician's ubiquitous urn), 95% of which are red, we are in a position to reason as follows:

1^* For any property p, at least α of n-fold samples exhibit a proportion that matches the population
2^* S is an n-fold sample from this urn
$\Longrightarrow$ [with probability $\geq \alpha$]
3^* S matches the population
4^* S has a proportion of .95 red balls

5^* The proportion of red balls in this urn lies in the interval $[.95-\varepsilon, .95+\varepsilon]$[15]

The move from 1^* and 2^* to 3^* is itself a direct inference, on the understanding that S is an epistemically random sample with respect to matching the population. The major premise is underwritten by Bernoulli's theorem. The move from 3^* and 4^* to 5^* incorporates the information regarding the sample proportion and the definition of matching. But 5^* is not quite the simple statistical statement we are accustomed to dealing with: rather, it states that the proportion of red balls in the urn lies in the interval $[.95-\varepsilon, .95+\varepsilon]$. Provided that ε is small, however, the lower boundary of this interval is still a healthy majority.

We can now extend the argument to predictive inference regarding an as yet unsampled ball from the urn:

6^* *a* is a ball from this urn
$\Longrightarrow$ [with probability in the interval $[.95-\varepsilon, .95+\varepsilon]$]
7^* *a* is red

There is no use caviling at 1^*, which is a mathematical truism. From 2^* and 4^*, which state the size and composition of our sample, and 6^* (which merely identifies *a*), we may draw a conclusion regarding an as-yet-unexamined member of the population with a reasonably high level of confidence (again, modulo the assumption of epistemic randomness). And by increasing the size of the sample, we can render the interval arbitrarily small without reducing the confidence level. Hence, an increase in sample size will allow us to take the sample proportion as an arbitrarily good estimate for the proportion of red balls in the urn.

While 1* in the argument above is unquestionably true and indeed mathematically provable, the question of epistemic access to this truth is more vexed. There is no way to construe Bernoulli's theorem as it must be used for this inference as an epistemic principle belonging at the metalevel, although it would have been very attractive to do so. Given our argument in Chapter 4 that metaprinciples need to be vindicable in principle but that the metaregress is stoppable even if the subject himself cannot provide the vindication, it would have been a more welcome conclusion if we could have placed Bernoulli at the metalevel and concluded that only epistemologists need know anything about him. But the form of inductive argument just presented is available only to those capable of understanding 1* sufficiently well to know it directly as an *object-level* foundational premise. From the perspective of providing a justificatory inductive argument available not only to the "common man" but also to those ubiquitous animals and young children that hover in the back of an epistemologist's mind (even if he is a committed *a priorist*), this is somewhat worrisome.

But in point of fact, it may not be necessary to know the detailed mathematics undergirding Bernoulli's theorem to have a sufficient conceptual grasp of it to underwrite many of our common inductions. Consider a simple example: if we imagine a population of three members, two A's and one B, and if we imagine first taking a sample of one from the population, two of the three samples we can get will be as representative as any one-member sample can be. Moving to twofold samples (with replacement) gives us 5/6 samples that are either all A or half of each, not a bad fit. Among threefold samples, nearly half (4/9) will be a *perfect* match and only a handful are far off. It is fairly easy to extrapolate this to a sample of four, five, and so forth and to see that, the larger the sample gets, the more heavily the representative possible samples outweigh the non-representative ones, so that the very large logically possible samples will be almost entirely

representative. Although we can conceive of a world in which the single B is drawn every time, even in a large sample, there are far more logically possible ways to get a large sample consisting either of all A's or of a ratio of approximately 2 to 1 A's over B's.

Despite the difficulties attending the attempt to determine what counts as justificatory access to premise 1*, this solution to the problem of induction is of more than academic interest. *Prima facie*, it is a cogent response to Hume's challenge. Hume himself grants that we have experience of bread's nourishing us and of the sun's rising. If we may take our experience to be a sample, then it appears that we possess all the tools necessary to make a rational defense of everyday extrapolations against Humean skepticism. But philosophical battles are not so easily won. Virtually every aspect of the argument just presented has been called into question. To these objections we now turn.

Linear attrition

A surprisingly common objection to this use of direct inference is that it reflects merely a linear elimination of alternative hypotheses regarding the composition of the population but offers no information regarding unexamined cases. A. J. Ayer suggests this argument in his description of a sampling experiment without replacement:

> If there are a hundred marbles in a bag and ninety-nine of them are drawn and found to be green, we have strong confirmation for the hypothesis that all of the marbles in the bag are green, but we have equally strong confirmation for the hypothesis that 99 per cent of the marbles are green and 1 per cent some other colour.[16]

In other words, drawing 99 balls from this bag gives us information precisely regarding the 99 balls in question, nothing more, nothing less. No matter how extensive our sample, the veil of ignorance always stands between us and the unsampled remainder.

John Foster, in his excellent book on Ayer, faithfully reproduces this criticism and explicates it with great clarity. Mathematical arguments designed to show that favorable instances increase the probability of a generalization, says Foster, reflect

> merely the trivial fact that, with each new favourable instance, there are fewer cases left in which the generalization could fail. The point is seen most easily if we focus on the example of drawing balls from a bag. Let us assume, for simplicity, that we already know that there are ten balls in all and that each is either black or white. When we have drawn one ball and found it to be black, we have more reason to believe that all the balls are black, simply because there are now only

> nine possible counter-instances remaining. ... This has nothing to do with induction, since it does not involve using the examined cases as a basis for predicting the properties of the unexamined cases. It tells us that the probability that *all* the balls are black increases, as the number of black balls drawn increases, but not that the probability that the *next* ball drawn will be black increases, as this number increases. Thus it does not tell us that, having drawn nine balls, we are entitled to be more confident about the colour of the tenth ball than when we first began the experiment.[17]

The implication of this argument is that a direct inference does not do the job required. Ayer and Foster are concerned that the data conveyed by a sample speak only for themselves and not for the unexamined cases. And on this reasoning, the promised inductive probabilities from sample data are a will-o'-the-wisp.

All of this is half right. Surely, an inductive argument is of no value unless it gives us, on the basis of examined cases, a justification for our beliefs regarding unexamined ones. But as an explication of the mathematical rationale for direct inference, the thesis of linear attrition is demonstrably wrong. To see this, we need only shift to sampling from Ayer's bag with replacement – creating, in effect, an indefinitely large population with a fixed frequency. No finite sample with replacement, no matter how large, ever amounts to a measurable fraction of this population. Yet using Bernoulli's theorem it is simple to specify a sample size large enough to yield as high a confidence as one likes that the true population value lies within an arbitrarily small (but non-degenerate) real interval around the sample proportion.

The thesis of linear attrition resembles an intuitively plausible error to which many beginning students of statistics are prone, namely, the mistake of thinking that the value of information in a sample is a function of the proportion of the population sampled. In fact, the relative proportion of the population sampled is not a significant factor in these sorts of estimation problems. It is the sheer amount of data, not the percentage of the possible data, that determines the level of confidence and margins of error. This consideration sheds some light on the worry raised by Peter Caws:

> Scientific observations have been made with some accuracy for perhaps 5,000 years; they have been made in quantity and variety only for about 500 years, if as long. ... [T]hese periods represent an almost infinitesimal fraction of the whole life of the universe. Further, all these observations have been made within a very thin spherical shell surrounding one planet of a small star. It may be that an animal species thus restricted in time and space has, in fact, succeeded in discovering the principles according to which the cosmos operates, but if it were not for the fact that we ourselves are members of this species, we should find the *a priori* probability of this rather small.[18]

Caws is certainly right to doubt whether every present regularity may properly be extrapolated into the misty past. But the grounds of such doubt have to do with our concrete evidence for differing conditions in the past rather than with the small fraction of time in which we have sampled the aeons. When we have no reason to believe conditions were relevantly different – as in the case, say, of certain geological processes – we may quite rightly extrapolate backwards across periods many orders of magnitude greater than those enclosing our observations.

Randomness, fairness, and representative samples

Or may we? There is a sharp division of opinion on the issue of randomness, and the defense of direct inference sketched above takes its stand on what is, admittedly, the more thinly populated side of the line. For four decades Henry Kyburg has stood almost *solus contra mundum* in his insistence that randomness is epistemic, that it is a primitive notion rather than something to be defined in terms of probability, and that in conjunction with statistical data it yields probabilities without "fair sampling" constraints.[19] We think he is right; and an examination of the problems generated by the standard definition of randomness indicates why this understanding of randomness is so important.

The standard statistical approach defines "randomness" in terms of equiprobability: a selection of an n-fold set from a population is random just in case every n-fold set is as likely to have been drawn from that population as any other.[20]

"But surely," runs the argument, "it is incumbent upon the defenders of direct inference to make some sort of defense of the claim that the sample selected was no more likely to be chosen than any other. The assumption is not generally true. Elementary textbooks are replete with examples of bias in sampling. To assume without argument that one's sample is unbiased is more than imprudent: in effect, it attempts to manufacture valuable knowledge out of sheer ignorance."

No other single criticism is more widely canvassed or more highly regarded in the literature. Apropos of an example involving a sample of marbles selected one each from 1000 bags, each of which contains 900 red and 100 white balls, Ernest Nagel urges that while Bernoulli's theorem

> does specify the probability with which a combination belonging to M [the set of all possible 1000-fold samples, one from each bag] contains approximately 900 red marbles, it yields no information whatever concerning the proportion of combinations satisfying this statistical condition that may be *actually selected* from the 1000 bags – unless, once more, the independent factual assumption is introduced that the ratio in which such combinations are *actually selected* is the same as the ratio of such combinations in the set M of all *logically possible* combinations.[21]

Without a special assumption of "fair sampling," we are vulnerable to the possibility that some samples may be much more likely to be selected than others; and perhaps the ones most likely to be selected are highly unrepresentative. Isaac Levi explicitly urges the need for such restrictions on direct inference in his critique of Kyburg.

> Suppose X knows that 90% of the Swedes living in 1975 are Protestants and that Petersen is such a Swede. Imagine that X knows nothing else about Petersen. On Kyburg's view, X should assign a degree of credence equal to .9 to the hypothesis that Petersen is a Protestant.
>
> I see no compelling reason why rational X should be obliged to make a credence judgment of that sort on the basis of the knowledge given. X does not know whether the way in which Petersen came to be selected for presentation to him is or is not in some way biased in favor of selecting Swedish Catholics with a statistical probability, or chance, different from the frequency with which Catholics appear in the Swedish population as a whole. . . .
>
> For those who take chance seriously, in order for X to be justified in assigning a degree of credence equal to .9 to the hypothesis that Petersen is a Protestant on the basis of direct inference alone, X should know that Petersen has been selected from the Swedish population according to some procedure F and also know that the chance of obtaining a Protestant on selecting a Swede according to procedure F is equal to the percentage of Swedes who are Protestants.[22]

Here is a pretty puzzle. We set out initially in search of a form of inference that would supply something we lacked: a rationally defensible ascription of probabilities to contingent claims on the basis of information that did not entail those claims. If we are required for the completion of this task to have in hand already the probability that this particular sample would be drawn (and indeed an identical probability for the drawing of each other possible sample), or information on the "chance" of obtaining a given sort of individual from the population (above and beyond frequency information), then the way is blocked. Direct inference is impaled on the empirical horn of Hume's dilemma.

In fact, if this criticism works at all, it applies more broadly than Levi indicates here. To infer from known black ravens that the next raven will be black, one must use direct inference *twice* – once to infer by Bernoulli's theorem the makeup of the population from the sample taken, and again to infer that the next item taken from the population will have a given property. Levi is, in this context, discussing only the second move; but if fair sampling constraints are required there, then they will apply *a fortiori* to the first. It is a natural extension of Levi's requirement to insist that we know a sample has been drawn by a fair method before we use it as a guide to the population.

The first step toward answering this criticism is to distinguish a "fair" sample from a "representative" one. Fair samples are drawn by a process that gives an equal probability to the selection of each possible sample of that size; a *representative* sample exhibits the property of interest in approximately the same proportion as the overall population from which the sample is drawn. To insist on a guarantee that the sample be representative in this sense is to demand something that turns induction back into deduction, for if we are certain that the sample is representative, we know *eo ipso* approximately what the population proportion is.

If a guarantee of representativeness is too much to demand, however, fairness seems at first blush to be a just requirement. We should like to avoid biased (i.e., unrepresentative) samples; and since most of the possible large samples are representative, a selection method that gives each such sample equal probability of being selected yields an agreeably high probability that a given sample is representative or that a given individual is typical of the population. Fair sampling will on occasion turn up samples that are wildly unrepresentative or individuals who are atypical. But constraints of the sort outlined by Levi, if we could be sure they held good, would assure us that in the long run these biased samples will make up only a small proportion of the total set of samples.

Here again, however, the road to an *a priori* justification of induction appears closed. For under the demands of fair sampling, we cannot rely on induction unless we know that each possible sample was equally likely to be chosen. And that is itself a contingent claim about matters that transcend our observational data and stands, therefore, in need of non-deductive justification. An infinite metaregress looms.

But the appeal of fairness constraints is illusory. Levi requires that X know Petersen has been selected by a method F, where the chance that F would obtain a Protestant from the population of Swedes is equal to the percentage of Swedes who are Protestants. Then, given the information that 90% of Swedes are Protestants, Levi allows that X could give a .9 credibility to the claim that Peterson is a Protestant, since the chance that a Protestant would be selected by method F is .9. But what does 'chance' mean here? If, as one must assume (especially given Levi's insistence on "taking chance seriously"), it means that 90% of the actual or possible applications of F would result in the selection of a Protestant from among Swedes, this reduces the problem to another direct inference, this one about instances of F rather than about Swedes. Knowing that 90% of all applications of F select Protestants could be helpful to us only if, on that basis, we were willing to repose a .9 degree of confidence in the claim that a given application of F (the one that produced Petersen) would produce a Protestant. But if this sort of answer were satisfactory there would have been no need to appeal to F in the first place. We could have placed a .9 confidence in the claim that Peterson is a Protestant based on the information that 90% of Swedes are Protestants. Similarly, we may place a high degree of

confidence in the proposition that a large sample is representative on the basis of Bernoulli's theorem. So Levi has not after all succeeded in replacing Kyburgian direct inference with a separate fairness requirement.[23]

Contrary to common wisdom, an assumption of fairness is *not* necessary for the epistemic legitimacy of direct inference. The demand for "objective fairness" is an empirical requirement; it has no bearing on the rationality of the inference either from sample to population or from population to individual. What is required instead is the condition that, relative to what we know, there be nothing about this particular sample that makes it less likely to be typical of the population than any other sample of the same size. This is the heart of the concept of *epistemic* randomness. Rather than locating randomness in the world as some real property of an object or a process, we locate it in the nature of our evidence about the object or sample. If we have positive knowledge, even if it is somewhat vague and indirect, that gives us actual reason to doubt that the sample or object is typical, rationality compels us to take this evidence into account. In the absence of such knowledge, the sample is to be treated simply as one of the logically possible samples of the population.

For this reason, when we speak of the individual in the minor premise of the direct inference as "a random member of the population with respect to" the property in question,[24] we are using a convenient manner of speaking. For on the theory we are defending, "randomness" is not literally a property of the member of the population but rather a property of our knowledge about the member of the population.[25] And in the presence of such epistemic randomness, no further concrete assumption of fairness in the method of sampling is required.

Even some critics of direct inference have recognized the justice of this point. Wisdom, for example, points out that it accords well with practical statistical work.

> We know in practical affairs that we must take random samples. But this is because we utilise existing knowledge. If we know of some circumstance that would influence a sample, we must look for a sample that would be uninfluenced by it. ... Now all this is only to say that *we avoid using a sample that is influenced in a known way.* ... If we demand that they should be random in some further sense, it is either a demand for knowledge of 'matching' or for additional knowledge about the influences that might affect the sample – the one would render statistical inference superfluous, the other is worthy in the interests of efficiency but does not come into conflict with Williams' argument. After all, *probability is used when all available knowledge has been taken account of and found insufficient*.[26]

An example makes this plain. Every Friday afternoon at 3:30 p.m. sharp, Professor Maxwell emerges from his office, strides down the hall to the

freshly stocked vending machine, inserts the appropriate amount of coins of the realm, and punches the button for a Coke. Because of the way the machine is designed, he will of course get the can resting at the bottom of the column: it is *that* can, no other, that will emerge. Yet given the information that one of the fifty cans in the vertical column is a Mello Yello and the other forty-nine are Coke, he is still justified in placing the probability that he will get a Coke at 98%. True, Maxwell is not equally likely to get any of the various cans stacked within the machine: his selection is not fair. But the Mello Yello is, on his information, a random member of the stack of cans with respect to position. Consequently, the can at the bottom is, on his information, a random element of the stack with respect to being a Coke.

The contrary intuition that demands fairness depends on what we might call a Cartesian worry rather than a Humean one: it conflates the presence of possibilities with the absence of probabilities.[27] If Maxwell sees that the machine has just been stocked by Damon, a resentful former logic student, he may harbor reasonable doubts that the can at the bottom is a Coke; it may not be a random member of the stack with respect to that property. (It may be a random member of the set of objects deliberately placed in someone's path by a practical joker intent on upsetting his victim's expectations – a set in which the frequency of anticipated outcomes is rather different!) But in the absence of some definite contrary evidence, the mere *possibility* that some can or other of the fifty might have been chosen deliberately to be placed at the bottom does not, in itself, provide information that changes the *probabilities* obtained by direct inference. And the fact that possibilities do not eliminate probabilities is a point that Descartes himself, for all his skeptical arguments, recognized very clearly.

The same considerations apply, *mutatis mutandis*, to sampling. The possibility that we might be sampling unfairly, like the logical possibility that Maxwell's nemesis has maliciously stacked the machine to trick him, cannot be eliminated *a priori*. But in the absence of concrete evidence that, e.g., places the about-to-be-selected sample in a different and more appropriate reference class, mere possibilities should not affect our evaluation of epistemic probabilities.

There are, of course, places where it is quite correct to demand certainty – to take a "Cartesian worry" very seriously indeed. It has been one of the burdens of this book to argue that *a priori* knowledge is infallible and that real epistemic principles must be knowable in this fashion. But here we are talking about object-level knowledge of empirical matters. The question at issue is whether it is necessary to eliminate the mere possibility of bias in order to be rational in making an inductive inference. But to put the question thus baldly is to make the answer evident. If we are not to beg the question against the possibility of rational non-deductive inference we must not demand *ab initio* that non-deductive inference be turned into deductive inference. And when there are (let us say) 99 black balls and one white ball

in the urn, it is always *possible* that we will choose the one white. Even if no one sees Damon tampering with the Coke machine, it is possible that he has been at work and that Maxwell will get the one Mello Yello among all the Cokes. But what do these possibilities say about rationality? Even if we could have guaranteed information about the makeup of the population, our selection in any given instance must always be of one option rather than another. And it will always be logically possible – even if we know nothing supporting such a supposition – that the selection mechanism is biased towards one part of the population. Even when we have positive evidence that a mechanism is not biased, we cannot eliminate the bare possibility. But if the rationality of the proportional syllogism is evident *a priori*, as we believe it is, then it must be possible to apply it rationally. And any attempt to apply it will require, at some point, a resort to epistemic randomness – an appeal to the fact that we have no reason to treat *this* object, *this* sample, or *this* application of a process differently from any other.

Appearances notwithstanding, this is not a retreat to the old principle of indifference; nor is it vulnerable to the charge, to which some advocates of that principle have exposed themselves, that it manufactures knowledge out of ignorance.[28] Indifference assigns equal probabilities to each element of a set on the basis of symmetry considerations, and a drawing method from that set is baptized "random" in terms of that assignment. On the account advocated here, by contrast, randomness is not parasitic on probability. Rather, *epistemic* randomness is fundamental to probability. To say that *a* is a random member of class F with respect to having property G, relative to my corpus of knowledge K, does invoke symmetry considerations. But when combined with knowledge of the frequency of G's among the F's, epistemic symmetry yields probabilities that reflect this relevant empirical information rather than reflecting hunches, linguistic symmetries or preconceived predicate widths. It is a consequence of this view that, in situations of complete ignorance regarding the proportion of F's that are G's, symmetry by itself yields no useful probability information: the probability is simply the maximally uninformative interval [0, 1].[29] This is an intuitively gratifying result. Epistemic symmetry conjoined with ignorance yields ignorance; conjoined with knowledge, it yields epistemic probabilities that reflect both the symmetry of information regarding set members and the knowledge we actually have about property distribution within the set.

Success versus rationality

The foregoing defense of randomness as a basis for assigning probabilities raises a fresh difficulty. The sort of "probability" that can be gotten from randomness and statistical information regarding a reference class is relativized, in the very definition of 'randomness,' to the state of our knowledge; and this strikes some critics as too much of a retreat from the goal of arriving at true beliefs. As a consequence, so runs the objection, any defense

of induction predicated on epistemic probability fails to address the true problem – the problem of future success.

This criticism recalls our reconstructed version of Hume's skeptical challenge: "Granted that these premises are true and that the conclusion is linked to them by a direct inference; why should that fact make the conclusion probable for me, in a sense that commends it to me if I prefer truth to falsehood?" By analogy with the natural answer regarding deductive inference, it would be at least *prima facie* satisfying to answer that direct inference guarantees a high proportion of future successes. But direct inference offers no such guarantee. Hao Wang puts the challenge succinctly when he notes that on an epistemic interpretation of probability

> we shall at no stage be able to pass from a certain frequency being overwhelmingly probable to it being overwhelmingly frequent. That is to say, on any non-frequency interpretation we have no guarantee that on the whole and in the long run the more probable alternative is the one that is more often realized.[30]

And again, criticizing Williams's *a priori* interpretation of probability, Wang asks:

> [W]hat guarantees induction to lead us more often to success than to disappointment, – granted that we can justify inductive generalizations with high probability on some *a priori* ground? . . . [A] principle of induction which might always lead to disappointment does not seem to be what is wanted. . . . [T]he conclusions reached in such fashion need not guarantee success, on the whole and in the long run, of our actions guided by them as predictions. In granting that we know *a priori* that a large sample very probably has nearly the same composition as the whole population, we must not forget that here what are known to be more probable need not be those which are on the whole and in the long run more often realized.[31]

Predictably, this line of criticism is advanced most vigorously by those who insist that both the definition of probability and the legitimacy of induction are bound up inextricably with contingent claims about the nature of the physical world. Nagel makes it clear that what makes Williams's justification of induction unacceptable to him is precisely this failure to guarantee success.

> For without the assumption, sometimes warranted by the facts and sometimes not, that a given method of sampling a population would actually select all samples of a specified size with roughly the same relative frequency, arithmetic can not assure us that we are bound to

> uncover more samples approximately matching the population than samples that do not.[32]

Why should such a "guarantee" or an "assurance" seem a compelling requirement for the justification of induction? Russell, in his defense of a finite frequency interpretation of probability, offers a clue. If we are obliged to admit (as Russell agrees we are) that the improbable may happen, then a probability claim that is not interpreted as a frequency

> tells us nothing about the course of nature. If this view is adopted, the inductive principle may be valid, and yet every inference made in accordance with it *may* turn out to be false; this is improbable, but not impossible. Consequently, a world in which induction is true is empirically indistinguishable from one in which it is false. It follows that there can never be any evidence for or against the principle, and that it cannot help us to infer what will happen. If the principle is to serve its purpose, we must interpret "probable" as meaning "what in fact usually happens"; that is to say, we must interpret a probability as a frequency.[33]

But the moral drawn here confuses success with rationality, an intrinsic with an extrinsic connection to truth. What Russell means by a world in which induction is "true" is, apparently, one in which inductive reasoning works well. Since it might turn out that all of our samples are unrepresentative, our extrapolations from them might all be hopelessly wide of the mark. This is, however, a reversion to what we have called the "Cartesian worry." It is possible to get a large but unrepresentative sample, just as it is possible to draw the one black ball from an urn of a million, 999,999 of which are white. But it would be irrational to expect this, given no further relevant information; and it is equally irrational to expect our samples to be unrepresentative and our inductions, in consequence, unsuccessful.

This confusion underlies Russell's complaint that such a principle "cannot help us to infer what will happen." If we demand a guarantee of success, or at any rate a guarantee of a high frequency of future successes, then we are indeed out of luck: that sort of "help" is not forthcoming. No amount of reasoning will turn contingent propositions into necessary ones. But rationality requires both less and more than this: less, because it is logically possible that a rational policy of non-demonstrative inference may always lead us astray; and more, because no accidental string of successes can in and of itself establish a policy of inference as rational.

Russell's argument is, in fact, vulnerable to the same response that undermines Levi's insistence on "fair sampling" constraints. A guarantee of a high proportion of successes is not only unavailable but would be useless without a subsequent appeal to unvarnished direct inference.[34] This is not merely because in the long run we are all dead: it applies even to an ironclad

guarantee that 99% of all of the inductions we make in the next year will be true. For in applications, it is always *this* induction, this particular instance, that is of importance. Even if it were granted that the proportion of successes among our inductions in the next year is .99 and that this application of inductive methodology is, given our present evidence, a random member of the class of those inductions with respect to its success, why should these facts confer any particular epistemic credibility upon the notion that this induction will be successful?

As we saw in Chapter 2, Michael Friedman raises the problem of a connection to truth in terms of proportions of successes. He demands that a method of reasoning be "reliable" in the sense that it yields success (i.e. true conclusions) in a high proportion as compared to the set of actual and physically possible applications of the method.[35] Here a problem of epistemic circularity arises for any attempt to derive the "reliability" of those methods from extant scientific theories, for those theories have nothing to commend them except that we have arrived at them by our non-deductive methods.

Moreover, Friedman's modal concept of objective chance falters, as do all demands that induction be guaranteed to be "mostly successful," on a problem of selective skepticism; the apostles of success must invoke the very reasoning they officially spurn. Why should a high proportion of successes among applications of our methods in any world, even our own, lead us to trust *this* application? It can only be because it is rational to expect success in this case based upon the knowledge that most applications of the method are successful. But again, if that response is legitimate for *methods of reasoning*, it is also good for sunsets, ravens, watermelons, and Swedes. The rationality of direct inference is so fundamental that the proposed alternative is epistemically valueless without a covert admission that direct inference is rational.

Once we have seen this, we are freed from the trap of thinking that a proper justification of induction must necessitate future success. There is no need to place new empirical information about the *successfulness* of induction at the metalevel in order to defend induction. The correct response to the modern Humean challenge regarding probabilities is to distinguish it from anxiety over bare possibilities and, having done so, to point out the way in which direct inference is underwritten by the symmetry of epistemically equivalent alternatives with respect to concrete frequency data. That symmetry offers no binding promises with respect to the future, no elimination of residual possibilities of failure. But our probabilistic extrapolations are apt to fail *only* if our samples have been unrepresentative; and despair over this bare possibility is, at bottom, an instance of the same fallacy that drives the credulous to purchase lottery tickets because of the *possibility* of winning. To see this fixation on possibilities aright is to understand the legitimacy of direct inference and to recognize that the probabilities it affords us are, in every sense of the term, rational.

Sampling the future: The modal barrier

Granting that the rationality of direct inference is logically independent of its record of successes, it is subject to what appears at first sight to be a severe limitation: it applies only to the population *from which we are sampling*, and that population often seems much more restricted than the scope of our conclusions. C. D. Broad raises this consideration to cast doubt on any approach to the problem of induction that takes its cue from observed samples, both because of our "restricted area of observation in space" and because of the "distinction of past and future cases" – by which he means quite simply that the probability of our having met any future crow is zero.[36] It is impossible to sample the future. Wisdom picks up on Broad's criticism to supply a vivid image of the modal barrier that apparently blocks the use of direct inference from the past and present to the future:

> [I]f some balls in an urn were sewn into a pocket, we could not get a fair sample – or rather we could not get a sample at all. Likewise the 'iron curtain' between the present and the future invalidates inductive extrapolation about the composition of things behind the curtain – we cannot sample them from this side.[37]

This objection has a plausible ring, but it proves extraordinarily difficult to give a detailed explanation of just why the modal barrier should block direct inferences. Ayer grants as an arithmetical truism that an omniscient being who made every possible selection precisely once would necessarily find that most of his samples were typical; but, he argues, this fact is of no value for our own inferences.

> It hardly needs saying, however, that we are not in this position. . . . So far from its being the case that we are as likely to make any one selection as any other, there is a vast number of selections, indeed in most instances the large majority, that it is impossible for us to make. Our samples are drawn from a tiny section of the universe during a very short period of time. And even this minute portion of the whole four-dimensional continuum is not one that we can examine very thoroughly.[38]

To extricate ourselves from this predicament, says Ayer, we require

> two quite strong empirical assumptions. They are first that the composition of our selections, the state of affairs which we observe and record, reflects the composition of all of the selections which are available to us, that is to say, all the states of affairs which we could observe if we took enough trouble; and secondly that the distribution

> of properties in the spatio-temporal region which is accessible to us reflects their distribution in the continuum as a whole.[39]

He is prepared to grant the first assumption, provided that we have taken some precautions to vary our samples and test our hypotheses under different conditions to safeguard against bias. But the second one he finds deeply problematic. The problem is not just that we are intuitively disinclined to extrapolate our local sample billions of years into the future or billions of light-years across the visible universe. That problem can be resolved by restricting the field of our conjectures to our local cosmic neighborhood and the relatively near future, and such a restriction may guarantee that our sample is typical of the local region of spacetime. If we approach the matter in this fashion, then

> we can be certain, and that without making any further assumptions, that in many cases the percentages with which the characters for which we are sampling will be distributed among [the populations in which we are interested] will not be very different at the end of the future period from what they are now. This will be true in all those cases in which we have built up such a backlog of instances that they are bound to swamp the new instances, however deviant these may be. But this conclusion is of no value to us. For we are interested in the maintenance of a percentage only in so far as it affects the new instances. We do not want to be assured that even if these instances are deviant the final result will be much the same. If we make the time short enough, we know this anyway. We want to be assured that the new instances will not be deviant. But for this we do require a non-trivial assumption of uniformity.[40]

Ayer's adroit exposition almost succeeds in concealing the fact that he has smuggled in the thesis of linear attrition once again. The problem arises *not* because the unsampled instances are future, but rather because they are unsampled, and we want to be assured that the unsampled instances are not deviant. "New instances," then, are the ones about which we have no information, and if this objection works at all, it will work regardless of their temporal position. The modal barrier is simply the veil of ignorance seen from a particular point of view.

This analysis of the objection casts doubt on Ayer's distinction between the two assumptions he thinks we need. If we are going to be worried about the unrepresentativeness of our sample regarding the far reaches of spacetime on the grounds that those far reaches may be deviant, then why not also be worried about unexamined ravens in the local wood at the dawn of the twenty-first century, since they may be deviant as well? That we have varied the conditions of our observations is no defense against this possibi-

lity, for we wish (following Ayer's example) to know not merely that our sample is representative of the whole spatiotemporally local population but that it is representative of the unexamined instances within that population. And however uniform our sample heretofore, we cannot eliminate what Wisdom calls

> the theoretical [problem] of making an inference about unexamined things in view of the possibility that the universe might play some trick that would wreck our best calculated expectations.[41]

Thus the thesis of linear attrition, and with it the modal barrier, are grounded in the worry about possibilities that we have already met; for the fear that the universe might "trick" us is plainly a reversion to Maxwell's apprehensions regarding Damon. Why, to use Wisdom's own analogy, should we believe that the balls sewn into a pocket in the bag are specially unrepresentative of the whole? To be sure, if we had some information to that effect, then epistemic randomness would be violated and we could not rationally use direct inference. But Wisdom leaves no doubt that fear of the bare *possibility* that our samples might be unrepresentative lies at the root of his inductive skepticism, for in his critique of Williams he explicitly repeats the objection:

> It is true that in the absence of knowledge of factors influencing a sample we rightly use that sample as a guide and that with such knowledge we rightly reject a sample. But here the position is that we do not know whether or not there is an influence at work and we think it possible there may be. In view of this doubt we cannot regard the sample as a guide that has the required statistical reliability.[42]

Wisdom appears, again, to be granting Williams's point, but he takes away with the left hand what he gives with the right. We think it "possible," he says, that there may be a biasing influence at work in our sample. The impression given on a cursory reading is that we think this proposition *plausible*, and that would involve positive knowledge raising the probability of a bias. But Wisdom's explicit contrast with the situation where we actually have knowledge of biasing factors makes such an interpretation untenable. He must be taken to mean that when it is *possible* that a sample is biased, we cannot regard it as a guide. Yet the contingent proposition that a sample has been biased by some "influence" is always *possibly* true. Hence, Wisdom's objection to drawing a conclusion about the future from past instances is not grounded in any special property of future balls nor in any special difficulty in gaining information about them from sample data but in a worry about bare possibilities that would apply to all inferences from sample to population.

Conclusion

Such is the moral of our extended examination of direct inference and the problem of induction. In case after case, the challenges reduce to the fundamental objection that the possibility of error has not been eliminated. The thesis of linear attrition, the demand for fairness constraints, the insistence on a guarantee of success, and despair of breaching the modal barrier are all variants on the same underlying theme: the fear "that the universe might play some trick" on us. To this as to all externalist objections there is in the final analysis only one answer, as old as Herodotus:

> There is nothing more profitable for a man than to take counsel with himself; for even if the event turns out contrary to one's hope, still one's decision was right, even though fortune has made it of no effect: whereas if a man acts contrary to good counsel, although by luck he gets what he had no right to expect, his decision was not any the less foolish.[43]

Notes

1 Internalism and the Collapse of the Gettier Problem

1 E. Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23, 1963, 121–23, reprinted in L. Pojman (ed.) *Theory of Knowledge: Classic and Contemporary Readings*, Belmont, CA: Wadsworth, 1994, pp. 134–36. All subsequent page references are to this reprinting.
2 B. Russell, *The Problems of Philosophy*, New York: Oxford University Press, 1959, pp. 131–32. The book was first published in 1912.
3 Ibid., p. 132.
4 R. Meyers and K. Stern, "Knowledge Without Paradox," *Journal of Philosophy* 70, 1973, pp. 147–60.
5 J. Dreher, "Evidence and Justified Belief," *Philosophical Studies* 25, 1974, pp. 435–39.
6 R. Shope, *The Analysis of Knowing*, Princeton, NJ: Princeton University Press, 1983, pp. 24–26.
7 "[I]f S is justified in believing P, and P entails Q, and S deduces Q from P and accepts Q as a result of this deduction, then S is justified in believing Q." Gettier, "Is Justified True Belief Knowledge?" p. 135.
8 Meyers and Stern, "Knowledge Without Paradox," pp. 157–58. We are indebted to Richard Fumerton for pointing out the importance of knowing the entailment relationship with certainty.
9 W. Alston, "The 'Challenge' of Externalism," in R. Schantz (ed.) *The Externalist Challenge*, New York: de Gruyter, 2004, p. 41.
10 P. Kitcher, "The Naturalists Return," *The Philosophical Review*, 1992, vol. 101, pp. 59–60.
11 A. I. Goldman, "A Causal Theory of Knowing," *Journal of Philosophy* 64, 1967, pp. 357–72. For example, Goldman imagines a person who infers that a mountain erupted in the past from the presence in the area of lava and portrays the subject as envisaging (based on his background knowledge) a process by which the eruption caused the presence of the lava. He argues that some such process as the subject envisages must have been the actual cause of the presence of the lava for the subject to have knowledge (pp. 361–63).
12 P. Unger, "An Analysis of Factual Knowledge," *Journal of Philosophy* 65, 1968, pp. 157–66.
13 A. Plantinga, *Warrant and Proper Function*, Oxford: Oxford University Press, 1993, pp. 36–37.
14 Ibid., p. 33.
15 Ibid.
16 Ibid., p. 35.
17 Ibid., p. 34.
18 Lest it be thought that, by referring to causation, we are capitulating to externalism after all, we hasten to add that, on our analysis, the belief that his barn-like experience is caused by a real barn is an implicit assumption on the part of the subject.
19 Gettier, "Is Justified True Belief Knowledge?" p. 135.

20 Shope, *The Analysis of Knowing*, p. 24. For the sake of convenience, we will be using Robert Shope's helpful summaries of various examples.
21 Plantinga, *Warrant and Proper Function*, p. 32.
22 Ibid.
23 See T. McGrew, *The Foundations of Knowledge*, Lanham, MD: Littlefield Adams Books, 1995, pp. 66–67, for a discussion of implicit reasons.
24 Shope makes it clear that this is the sort of evidence intended. See *The Analysis of Knowing*, p. 4, footnote 3.
25 The use of "therefore" in arguments may be informal. We will emphasize throughout this chapter specific premises necessary to make the arguments in question rational (premises that often turn out to be false), but the arguments may be enthymematic in other ways, as well.
26 Shope, *The Analysis of Knowing*, pp. 24–25.
27 Ibid., p. 25.
28 R. Feldman, "An Alleged Defect in Gettier Counter-Examples," *Australasian Journal of Philosophy*, 52, 1974, p. 69. See also R. Feldman, *Epistemology*, Upper Saddle River, NJ: Prentice Hall, 2003, pp. 31–33. Feldman here maintains that in some sense S's reason "depends on" a falsehood but still explicitly denies that the falsehood is a ground or premise of the conclusion.
29 See B. Mates, *Elementary Logic*, New York: Oxford University Press, 1972, p. 120.
30 We will return to the question of the representation of non-deductive reasoning in Chapter 7, where this issue turns out to be important for understanding the problem of induction.
31 Shope, *The Analysis of Knowing*, p. 25.
32 Ibid.
33 B. Skyrms, "The Explication of 'X knows that *p*'," *Journal of Philosophy* 64, 1967, p. 383ff.
34 Shope believes that all Gettier cases involve a "significant falsehood," although he denies that such a falsehood is necessarily any part of the subject's argument. See *The Analysis of Knowing*, p. 25.
35 For an explanation of evidential symmetry – epistemic randomness – see Chapter 7.
36 Shope, *The Analysis of Knowing*, p. 222.
37 J. G. Dees and J. A. Hart, "Paradox Regained: A Reply to Meyers and Stern," *The Journal of Philosophy* 71, 1974, pp. 369–70.
38 For an earlier treatment of a related concept, that of a pruned evidence tree, see McGrew, *Foundations*, pp. 46, 51–54.
39 Meyers and Stern, "Knowledge Without Paradox," p. 149. K. Lehrer, "Knowledge, Truth, and Evidence," *Analysis* 107, 1965, pp. 168–75. See especially pp. 170–71.
40 Meyers and Stern, "Knowledge Without Paradox," pp. 149–50. They actually say that the subject 'could' give a new argument, but the strengthening of this to "would" seems justified in the light of other statements they make.
41 Dees and Hart, "Paradox Regained," pp. 370–71.
42 For example, Richard Fumerton maintains that inferential justification requires that the subject know that his premises make his conclusion probable. See R. Fumerton, *Metaepistemology and Skepticism*, Lanham, MD: Rowman and Littlefield, 1995, pp. 36–37.
43 This entire discussion applies only to *premises* of S's argument. In some cases, used in defeasibility analyses of the Gettier problem, the subject would drop his conclusion were he to learn of a "defeater" present in the situation, but he is in no sense using the contradictory of the defeater as a premise. Many of these arise because the subject is inferring (or at least would be inclined to infer) the falsehood of the "defeating" statement, either inductively or deductively, from the truth of his conclusion. E.g. S might be inclined to infer from his belief that he does not have a disease – based on evidence such as the small incidence of the disease in his society, his being at low risk personally, etc. – that a laboratory has not written on a piece of paper that he has the disease. If S found out that

the lab did have a positive result written down, this would be evidence against the proposition that he was disease-free, for straightforward Bayesian reasons. However, at t S's belief that he is disease-free is based on entirely independent evidence, not on the falsehood of the proposition about the lab's paperwork. Our analysis entails that, even if the subject would drop his conclusion if he were to learn of such a "defeater," its mere existence does not actually defeat his justification and is, in fact, irrelevant to the question of whether he has knowledge.

44 H. Kornblith, "Beyond Foundationalism and the Coherence Theory," in H. Kornblith (ed.) *Naturalizing Epistemology*, Cambridge, MA: MIT Press, 1985, pp. 117–19.

45 See T. McGrew and L. McGrew, "Psychology for Armchair Philosophers," *Idealistic Studies* 28, 1998, pp. 147–57.

46 Dreher, "Evidence and Justified Belief," p. 436.

47 Meyers and Stern make this point when they say that a premise one wishes to use for a knowledge-granting inference need only be true, not certain. Meyers and Stern, "Knowledge Without Paradox," pp. 152–53.

48 Cf. McGrew, *Foundations*, pp. 50–51.

49 We shall have much more to say about the "no new empirical information" aspect of our position in Chapter 3.

50 Plantinga, as mentioned earlier, rejects the Russellian solution summarily, moving on rapidly to his own "proper function" explanation based on factors inaccessible to the subject. It is clear that Plantinga does not consider the Russellian analysis of the Gettier problem to capture the externalist implications he sees in Gettier. See *Warrant and Proper Function*, pp. 32–36.

51 An earlier version of this article appeared as Timothy and Lydia McGrew, "Internalism and the Collapse of the Gettier Problem," *Journal of Philosophical Research* 23, 1998, pp. 239–56.

2 The Connection to Truth

1 R. Foley, *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press, 1987, pp. 167–69. *Working Without a Net: A Study of Egocentric Epistemology*, Oxford: Oxford University Press, 1993, pp. 85–87.

2 S. Cohen, "Justification and Truth," *Philosophical Studies* 46, 1984, p. 279.

3 R. Fumerton, *Metaepistemology and Skepticism*, Lanham, MD: Rowman and Littlefield, 1995, pp. 192, 200–202.

4 L. BonJour, "Plantinga on Knowledge and Proper Function," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman and Littlefield, 1996, pp. 53–55.

5 Fumerton, *Metaepistemology and Skepticism*, p. 97.

6 A. I. Goldman, "The Internalist Conception of Justification," in P. French *et al.* (eds) *Midwest Studies in Philosophy V: Studies in Epistemology*, Minneapolis, MN: University of Minnesota Press, 1980, p. 32.

7 Ibid., pp. 38, 45.

8 B. Russell, *Human Knowledge: Its Scope and Limits*, New York: Simon and Schuster, 1948, p. 401.

9 M. Friedman, "Truth and Confirmation," in H. Kornblith (ed.) *Naturalized Epistemology*, Cambridge, MA: MIT Press, 1985, pp. 155–56.

10 D. Miller, *Critical Rationalism: A Restatement and Defense*, Chicago, IL: Open Court, 1994, p. 66.

11 Ibid.

12 BonJour, "Plantinga on Knowledge and Proper Function," pp. 53–55.

13 A. Plantinga, "*Respondeo*: *Ad* BonJour," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology*, p. 342. Also A. Plantinga, *Warranted Christian Belief*, Oxford: Oxford University Press, 2000, p. 175.

14 W. Alston, "Epistemic Circularity," in *Epistemic Justification*, Ithaca, NY: Cornell University Press, 1989, p. 326.
15 R. Audi, "Contemporary Foundationalism," in Louis Pojman (ed.) *The Theory of Knowledge: Classical and Contemporary Readings*, Belmont, CA: Wadsworth Publishing Co., 1998, p. 207. It is not clear exactly what significance Audi attaches to this connection to truth, since elsewhere he seems to disavow a need for actual reliability of belief sources. See R. Audi, *The Structure of Justification*, Cambridge: Cambridge University Press, 1993, pp. 314–22.
16 E. Fales, *A Defense of the Given*, Lanham, MD: Rowman and Littlefield, 1996, pp. 174–76.
17 T. McGrew, *The Foundations of Knowledge*, Lanham, MD: Littlefield Adams Books, 1995.
18 In our view, this is because they possess a referential relation to experiential states which render them true necessarily.
19 In Chapter 6 we will respond to the externalist *tu quoque* argument that *a priori* consistency proofs in deductive logic have the same property of epistemic circularity that internalists find objectionable in externalist defenses of contingent epistemic principles.
20 In saying this we are consciously contradicting L. Jonathan Cohen's multicriterial theory of probability as he lays it out in *The Probable and the Provable*, Oxford: Clarendon Press, 1977, Chapter 2. We agree with his contention that provability is a limiting case of probability, but unlike Cohen we do not believe that there is any contingent or singular relation of provability.
21 We are indebted to Richard Fumerton for emphasizing this point in private communication.
22 The idea is not idiosyncratic. See A. de Morgan, *Formal Logic, or the Calculus of Inference, Necessary and Probable*, London: Taylor and Walton, 1847, and particularly J. M. Keynes's *A Treatise on Probability*, London: Macmillan, 1963, pp. 133 ff.
23 It would be possible to give such a semantics entirely in terms of class ratios, and frequentists will insist that this is the only correct interpretation of probability even for epistemic purposes. In Chapter 7 we will elaborate our reasons for adopting an epistemic rather than a frequentist conception of probability.
24 This is the sum of the uncertainties of the premises. See E. W. Adams, *A Primer of Probability Logic*, Stanford, CA: CSLI Publications, 1998.
25 Keynes, *A Treatise on Probability*, pp. 3–4.
26 This terminology is not intended to imply that subjects can have beliefs only if they possess the abstract concepts of truth and falsity.
27 See L. BonJour, *Epistemology: Classic Problems and Contemporary Responses*, Lanham, MD: Rowman & Littlefield, 2002, pp. 236–37 for a similar point on the priority of the internalist approach and the metaregress created by externalism.
28 A. Plantinga, "*Respondeo*: *Ad* BonJour," p. 342; *Warranted Christian Belief*, p. 130, n. 23 and p. 221; *Warrant, the Current Debate*, Oxford: Oxford University Press, 1993, p. 100; cf. *Warrant and Proper Function*, Oxford: Oxford University Press, 1993, pp. 236–37.
29 A. Plantinga, *Warranted Christian Belief*, pp. 128–30.
30 Plantinga, "*Respondeo*: *Ad* BonJour," p. 341.
31 Ibid., p. 342.
32 Descartes appears to have had something like this distinction in mind in the second set of replies. See J. Cottingham *et al*., (eds and trans.) *The Philosophical Writings of Descartes*, Cambridge: Cambridge University Press, 1984, vol. 2, p. 104. But the interpretation of Descartes on whether it is possible to be wrong about truths clearly and distinctly perceived is sufficiently controversial that we do not intend to take a position here on the historical question.
33 See Chapter 5 for further discussion of intuition of *a priori* truths and of Plantinga on phenomenology and *a priori* knowledge.
34 Fumerton, *Metaepistemology and Skepticism*, pp. 204, 216–18.
35 Ibid., p. 221.
36 Ibid., p. 218.

37 Ibid., p. 215.
38 Ibid., p. 204.
39 Ibid., p. 194.
40 Friedman, "Truth and Confirmation," p. 156.
41 For some results linking IBE and Bayesian probability, see T. McGrew, "Confirmation, Heuristics, and Explanatory Reasoning," *British Journal for the Philosophy of Science* 54, 2003, pp. 553–67.
42 It is worth noting that the position outlined here is very different from John Pollock's position, which Fumerton discusses briefly (p. 193). Pollock's "conceptual analysis" is closely related to verificationism. The crucial concepts on his position are the concepts of empirical objects themselves, understood in terms of their justification conditions. See J. Pollock, *Knowledge and Justification*, Princeton, NJ: Princeton University Press, 1974. It is certainly impossible to avoid using the concepts which play a part in a particular hypothesis when one is evaluating the argument for that hypothesis. But the analytic approach we advocate concentrates on concepts involved in the form of the argument itself.
43 S. Cohen, "Justification and Truth," p. 291.
44 Fumerton, *Metaepistemology and Skepticism*, pp. 194–95.
45 Chisholm seems to do exactly this in one of his "epistemic principles" which states, "If S *seems to remember* having been F, and if it is epistemically in the clear for him that he remembers having been F, then it is *beyond reasonable doubt* for S that he remembers having been F." R. Chisholm, *Theory of Knowledge*, 3rd edn, Englewood Cliffs, NJ: Prentice Hall, 1989, p. 68.
46 Fumerton, *Metaepistemology and Skepticism*, pp. 108–11, 203.
47 Laurence BonJour has made the irenic suggestion that externalists and internalists are simply pursuing different interesting epistemological projects and has identified the externalist project as the attempt to discover which of our belief-forming methods are in fact reliable (L. BonJour and E. Sosa, *Epistemic Justification*, Oxford: Blackwell Publishing, 2003, pp. 36–37). We would respond that this is not an "externalist project" in the sense that internalists are *debarred* from pursuing it. But it is an empirical inquiry, *not* an epistemological one. The claim that it is part of *epistemology* marks one as, to at least some degree, sympathetic to externalism.
48 And in fact, such a procedure yields the metaregress we discuss in the following chapters.

3 Internalism, Externalism, and the Metaregress

1 See W. Alston, *Epistemic Justification*, Ithaca, NY: Cornell University Press, 1989, pp. 321–22; R. Nozick, *Philosophical Explanations*, Cambridge, MA: Belknap Press, 1981, p. 172; and A. I. Goldman, *Epistemology and Cognition*, Cambridge, MA: Harvard University Press, 1986, pp. 38–39 *et passim*. In each case the version of externalism advanced allows that high-level empirical beliefs may be justified (or fulfill a similar role in empirical knowledge to the role justification is supposed to play) even if a subject does not have an argument of the internalist sort for these beliefs, but the beliefs cannot be justified in the *absence* of the externalist connection. For our purposes, the differences among the various formulations are less important than their similarities.
2 And, perhaps an idealized, freeze-framed slice of time in which to introspect; though this may be obviated if the subject were to have sufficient intellectual clarity at a given moment.
3 Alston, *Epistemic Justification*, p. 342. We expand on the notion of an "upgraded" subject below and again in Chapter 4.
4 See K. Korcz, "Recent Work on the Basing Relation," *American Philosophical Quarterly* 34, 1997, pp. 171–91 for a useful survey.
5 Richard Swinburne advocates a position of this sort in *Epistemic Justification*, Oxford: Oxford University Press, 2001, pp. 132–33. For the account of causation he employs, see "The Irreducibility of Causation," *Dialectica* 51, 1997, pp. 79–92.

6 The lack of interest in the *sui generis* position may arise from a desire to derive some further consequences about how reasons guide diachronic belief formation from the analysis of basing, a motivation we do not share. See, for example, the detailed discussion by Robert Audi, "Belief, Reason, and Inference," in *The Structure of Justification*, Cambridge: Cambridge University Press, 1993.
7 See, A. Plantinga, *Warrant and Proper Function*, Oxford: Oxford University Press, 1993, pp. 17–19 and 183–84.
8 See, for example, the discussion in Plantinga's *Warrant and Proper Function*, pp. 172–73.
9 Although Plantinga's position cannot be described as garden-variety reliabilism, his conditions for warrant have a clear reliability component. See *Warrant and Proper Function*, pp. 17ff.
10 Aristotle, *Posterior Analytics*, Book I, chapter 3, 72b5–73a7.
11 W. Alston, *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press, 1993, pp. 3–4.
12 M. Bergmann, "Externalism and Skepticism," *The Philosophical Review* 109, 2000, pp. 159–94. The passage in question appears on 188–89.
13 Plantinga, *Warrant and Proper Function*, p. 172.
14 Perhaps even the occasional contradiction could be properly basic if believing it induced the subject to derive from it, by the well-known law of Duns Scotus, a great number of truths that he would not otherwise have believed.
15 L. BonJour, "Plantinga on Knowledge and Proper Function," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman and Littlefield, 1996, pp. 53–55. As we are using the term 'empirical' in the concept of armchair internalism, we mean to exclude new *factual* information from the metalevel. BonJour indicates that the rareness of oak trees is an "external" matter, yet he holds out hope that it might be determinable *a priori*; he admits that it will not work to investigate it empirically. It seems that his hope arises from the fact that he is considering "rareness" in relation to the total set of *possible* worlds. Yet on BonJour's own rationalist account of the *a priori*, such knowledge will be synthetic and factual, so it will be "empirical" in our sense (see L. BonJour, *In Defense of Pure Reason*, New York: Cambridge University Press, 1998). The crucial point is that he is treating reliability claims as metalevel, a position that armchair internalists reject.
16 In some respects this dilemma resembles one developed by Michael Bergmann in "A Dilemma for Internalism," in T. Crisp *et al*. (eds) *Knowledge and Reality: Essays in Honor of Alvin Plantinga*, Dordrecht, Netherlands: Springer, 2006, pp. 137–77. Available HTTP: <http://web.ics.purdue.edu/~bergmann/dilemma.htm> (accessed September 30, 2005).
17 See the extended discussion of such requirements in Chapter 4.
18 R. Fumerton, *Metaepistemology and Skepticism*, Lanham, MD: Rowman and Littlefield, 1995, p. 86.
19 Bergmann, "A Dilemma for Internalism," sec. 2.2.
20 Fumerton maintains that a subject whose premises strictly entail his conclusion must believe that the entailment holds in order to be justified at the object level. Fumerton's position is that there is no justificatory inference at the object level (as opposed to a mere causal connection) if there is no justified belief that the premises make the conclusion probable. "Inferential Internalism and the Presuppositions of Skeptical Arguments," in R. Schantz (ed.) *The Externalist Challenge*, New York: de Gruyter, 2004, pp. 165–66.
21 W. Alston, "Epistemic Circularity," in *Epistemic Justification*, p. 326.
22 Ibid., p. 328.
23 W. Alston, *The Reliability of Sense Perception*, p. 15.
24 A. Plantinga, *Warranted Christian Belief*, Oxford: Oxford University Press, 2000, p. 119.
25 There are forms of externalism, including of course Plantinga's own, that involve evaluating positive epistemic status without using the term 'justification.' An exact parallel of our definition of epistemic circularity applies to these theories. E.g. "Epistemic circularity

is present whenever *p* appears in the argument for *Wp*," where 'W' denotes "is warranted." The same is true of our definition of the metaregress, where the hierarchy would involve the iteration of a metalevel predicate other than 'J.'

26 Alston, "Epistemic Circularity," in *Epistemic Justification*, p. 345.

27 For more on epistemic circularity and other kinds of infinite metaregress, see T. McGrew and L. McGrew, "Level Connections in Epistemology," *American Philosophical Quarterly* 34, 1997, p. 87.

28 In fact, Alston, like Plantinga, seems inclined to apply a reliability requirement even to beliefs in non-contingent propositions, such as the axioms of deductive logic. Alston, *Reliability of Sense Perception*, p. 116. See next note.

29 E.g. A. Plantinga, "Internalism, Externalism, Defeaters and Arguments for Christian Belief," *Philosophia Christi*, series 2, 3:2, 2001, p. 390.

4 What's Wrong with Epistemic Circularity

1 Laurence BonJour makes a similar point in *Epistemology: Classic Problems and Contemporary Responses*, Lanham, MD: Rowman & Littlefield, 2002, p. 235.

2 W. Alston, *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press, 1993, p. 16. Cf. "Epistemic Circularity," in *Epistemic Justification*, pp. 330–31, 334–35, 348–49.

3 R. Fumerton, *Metaepistemology and Skepticism*, Lanham, MD: Rowman & Littlefield, 1995, p. 36.

4 Fumerton, *Metaepistemology and Skepticism*, pp. 81–82.

5 In personal correspondence, Fumerton has indicated that this latter idea is nearer to what he has in mind.

6 Fumerton, *Metaepistemology and Skepticism*, pp. 85–87. Cf. Fumerton on "primary" and "secondary" epistemic principles, pp. 104–5.

7 The phrase 'every hierarchy' reflects the fact that, for the claims which constitute the premises and conclusion of the argument at any given level, separate metajustifications may be required.

8 Alston, *Reliability of Sense Perception*, p. 125, n. 5.

9 A. Plantinga, *Warranted Christian Belief*, Oxford: Oxford University Press, 2000, p. 125. Cf. Plantinga, "*Respondeo*: *Ad* BonJour," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology*, Lanham, MD: Rowman & Littlefield, 1996, p. 342.

10 J. Watkins, *Science and Scepticism*, Princeton, NJ: Princeton University Press, 1984, pp. 258–59. We are adopting Watkins's term, although he is concerned only with Johnny Wideawake's object-level reasons and has specific ideas about his limitations to which we are not committed.

11 Alston, *Reliability of Sense Perception*, p. 17.

12 Alston, "Epistemic Circularity," in *Epistemic Justification*, p. 331. This proposition is numbered V in Alston's argument, a designation we have altered.

13 W. Alston, "Level Confusions in Epistemology," in *Epistemic Justification*, p. 170. Cf. p. 168. This point is also the burden of much of "Epistemic Circularity."

14 Alston, "Epistemic Circularity," in *Epistemic Justification*, pp. 342–43.

15 Since externalism is our target, we shall confine ourselves to arguing for SMP′. We shall not argue for – although we would advocate – the requirement that someone (such as Johnny Wideawake) be able to stop the metaregress for relevant statements that make specific reference to an actual subject's belief – e.g. "S's belief that *p* is the conclusion to an inductive argument."

16 It is not uncommon for epistemologists to use the phrase 'epistemic principles' to refer to principles like EP that state specific conditions for justification. Others besides Alston who do so are Roderick Chisholm, *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 1989, pp. 61ff, and Robert Audi, *Belief, Justification, and Knowledge*, Belmont, CA: Wadsworth, 1988, pp. 38, 49. Cf. Ralph Baergen's use of the phrase 'epistemic first principles' in expounding Thomas Reid (R. Baergen, *Contemporary Epistemology*, New

York: Harcourt, Brace, 1995, pp. 146–49). The point is not that these are all externalists – they are not – but rather that the phrase has currency with reference to statements that, from an externalist perspective, cannot be necessary truths.

17 See Alston, "Epistemic Circularity," *passim*.

18 T. McGrew and L. McGrew, "What's Wrong With Epistemic Circularity," *Dialogue* XXXIX, 2000, pp. 226–32.

19 Whenever we say that a proposition is "indefensible" or "cannot be defended," we mean to imply that it stands in need of a defense, whether that be at the object level in the form of evidence or at the metalevel in the form of a vindication. We shall argue in the next chapter that metafoundations do not stand in need of further vindication, just as object-level foundations do not stand in need of evidential support.

20 Of course we might pick up on the resulting regularity by an inductive procedure, as Reichenbach stresses. H. Reichenbach, *Experience and Prediction*, Chicago: University of Chicago Press, 1938, pp. 353–54. But under such circumstances, the claim that there was such a regularity would not be an epistemic principle but rather an empirical generalization, epistemically useful to us only insofar as it functioned as a premise at the Aristotelian level of the argument. The epistemic principle in question would be the inductive one.

21 The reference to "showing unequivocally" is aimed specifically at Alston's claim in "Epistemic Circularity" that epistemic circularity does not preclude "showing" that a belief is justified (pp. 333–34). But it becomes clear that such a "showing" will, on Alston's view, involve an argument at the metalevel that itself stands in need of vindication and that is justificatory only conditionally. This is not what epistemic principles are supposed to do, and it provides no help to one who wishes to avoid the epistemic anarchy just described.

22 Fumerton, *Metaepistemology and Skepticism*, p. 177. See pp. 177–79 for Fumerton's further discussion, which our analysis closely parallels, of the problems externalism creates for discriminating among epistemic principles.

23 B. Russell, *An Inquiry Into Meaning and Truth: The William James Lectures for 1940 Delivered at Harvard University*, London: Unwin Paperbacks, 1980, pp. 14–15.

24 R. Fumerton, "Plantinga, Warrant, and Christian Belief," *Philosophia Christi*, series 2, 3:2, 2001, 343–44.

25 M. Martin, *Atheism: A Philosophical Justification*, Philadelphia: Temple University Press, 1990, p. 276.

26 Cf. Plantinga's assertion that all principles about probability are contingent and that, therefore, anti-inductive inferences might be "justified" because they met the design plan on Alpha Centauri. *Warrant and Proper Function*, Oxford: Oxford University Press, 1993, pp. 172–73.

27 It is even simpler to make a similar objection to non-Plantingian forms of externalism, since one need not talk about design plans at all. Whatever truth-conducive process or truth-indicative feature the externalist demands for his version of "justification," it is an empirical question whether Linus's belief fulfills that requirement, and any attempt to say that it *does not* fulfill it (whereas some apparently more plausible belief does) sets off a metaregress.

28 Plantinga, *Warranted Christian Belief*, p. 346.

29 Plantinga does attack the argument for Christianity from the resurrection as probabilistically poor (*Warranted Christian Belief*, pp. 268ff). Moreover, he states explicitly that he does not know of *any* good arguments for Christian belief ("Internalism, Externalism, Defeaters and Arguments for Christian Belief," *Philosophia Christi*, series 2, 3:2, 2001, 398). On Plantinga's critique, see T. McGrew, "Has Plantinga Refuted the Historical Argument?" *Philosophia Christi*, series 2, 6:1, 2004, 7–26.

30 Our criticisms are not exclusively focused on properly basic beliefs. In the case of properly basic beliefs the absence of internal rationality is perhaps most evident, but beliefs formed by an anti-inductive "design plan" also are obviously lacking in internal rationality,

and in general, internal rationality is not required by externalism either for inferred or for foundational beliefs. Furthermore, the metaregress problem affects externalist principles concerning both inference and foundational belief formation.

31 Plantinga, *Warrant and Proper Function*, p. 231.
32 E.g. Plantinga, *Warrant and Proper Function*, Chapter 12, "Is Naturalism Irrational?" *passim*, and *Warranted Christian Belief*, pp. 363ff., *passim*.
33 Plantinga, *Warrant and Proper Function*, pp. 164–66.
34 Plantinga, *Warranted Christian Belief*, p. 363.
35 Alston, *Reliability of Sense Perception*, pp. 124–26.
36 Plantinga, *Warranted Christian Belief*, pp. 127–29.
37 Alston, *Reliability of Sense Perception*, p. 133
38 Ibid.
39 Ibid., p. 138.
40 Ibid., p. 139.
41 Ibid., p. 17.
42 Ibid., p. 138.
43 For a lovely early articulation of this point, presented as a rebuttal of a hypothetical pragmatist, see S. Barker, *Induction and Hypothesis*, Ithaca, NY: Cornell University Press 1957, pp. 17–18.
44 Fumerton, *Metaepistemology and Skepticism*, p. 221.
45 Portions of this chapter were originally published in two earlier articles: Timothy and Lydia McGrew, "Level Connections in Epistemology," *American Philosophical Quarterly* 34, 1997, 85–94 and Timothy and Lydia McGrew, "What's Wrong with Epistemic Circularity," *Dialogue* 39, 2000, 219–39.

5 Analytic *a priori* Knowledge

1 See T. McGrew, *The Foundations of Knowledge*, Lanham, MD: Littlefield Adams Books, 1995, and also "A Defense of Classical Foundationalism," in L. Pojman (ed.) *The Theory of Knowledge: Classical and Contemporary Readings*, 2nd edn, Belmont, CA: Wadsworth, 1998.
2 L. BonJour, *In Defense of Pure Reason*, New York: Cambridge University Press, 1998, p. 33 *et passim*.
3 This point crops up repeatedly in Russell's work on the foundations of mathematics. See, e.g., B. Russell and A. N. Whitehead, *Principia Mathematica to *56*, Cambridge: Cambridge University Press, 1962, p. v: "In mathematics, the greatest degree of self-evidence is usually not to be found quite at the beginning, but at some later point: hence the early deductions, until they reach this point, give reasons rather for believing the premises because true consequences follow from them, than for believing the consequences because they follow from the premises."
4 J. Locke, *An Essay Concerning Human Understanding*, vol. 2, A. C. Fraser (ed), New York: Dover, 1959, book IV, ch. 1, sec. 4, p. 169.
5 See the next chapter for a discussion of demonstration and certainty.
6 See, for example, R. I. Aaron, *John Locke*, 2nd edn, Oxford: Oxford University Press, 1955, pp. 224–25.
7 See below for our discussion of the "mere phenomenology" objection to this account of *a priori* knowledge.
8 B. Russell, *The Theory of Knowledge: The 1913 Manuscript*, published as *The Collected Papers of Bertrand Russell*, vol. 7, London: George Allen and Unwin, 1983, p. 157.
9 The notion of acquaintance went out of fashion in epistemology for roughly three decades beginning with Wilfrid Sellars's influential attack on it in *Science, Perception and Reality*, London: Routledge and Kegan Paul, 1963, pp. 131–32. BonJour endorsed this attack in *The Structure of Empirical Knowledge*, New York: Harvard University Press, 1985, ch. 5, but rejects it in "Toward a Defense of Empirical Foundationalism" in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman and Littlefield,

2001, pp. 21–38. For some work reviving and explicating the notion, see T. McGrew, *The Foundations of Knowledge*, R. Fumerton, *Metaepistemology and Skepticism*, Lanham, MD: Rowman and Littlefield, 1995, and R. Fumerton, "Classical Foundationalism," in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman and Littlefield, 2001, pp. 3–20.

10 C. I. Lewis, *An Analysis of Knowledge and Valuation*, LaSalle, IL: Open Court, 1946, p. 118.

11 BonJour, *In Defense of Pure Reason*, p. 39.

12 Ibid., pp. 106ff.

13 Ibid., pp. 32–33, 72–73.

14 H. Putnam, "The Meaning of 'Meaning'," in his *Philosophical Papers*, vol. 2, *Mind, Language and Reality*, Cambridge: Cambridge University Press, 1975, pp. 215–71; H. Putnam, "Meaning and Reference," in P. Moser (ed.) *Reality in Focus*, Englewood Cliffs, NJ: Prentice Hall, 1990, pp. 381–88.

15 Putnam, "Meaning and Reference," pp. 385–87.

16 T. and L. McGrew, "Psychology for Armchair Philosophers," *Idealistic Studies* 28, 1998, pp. 147–57. See especially pp. 152–55.

17 Putnam, "Meaning and Reference," pp. 382, 384, 388.

18 This is not merely an artifact of the way that we have phrased the objection. In addition to the references in the previous footnote to "Meaning and Reference," see J. Heil, "Hilary Putnam," in A. P. Martinich and D. Sosa (eds) *A Companion to Analytic Philosophy*, New York: Blackwell, 2001, p. 395: "If we think of the meanings of our terms as what fixes the extension of those terms . . . "

19 Putnam, "Meaning and Reference," p. 384.

20 See "Psychology for Armchair Philosophers" on opaque and transparent concepts. Note also that here we are using 'reference' in the looser, indirect sense in which one can refer to a horse at all. In our view one cannot refer *directly* to any extra-mental object but only to the immediate experiences with which one is directly acquainted.

21 For a vigorous defense of a Russellian theory of names, see R. Fumerton, "Russelling Causal Theories of Reference," in C. W. Savage and C. A. Anderson (eds) *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology*, Minneapolis: University of Minnesota Press, 1989, pp. 108–18.

22 Putnam, "Meaning and Reference," p. 388, n. 2; "The Meaning of 'Meaning'," pp. 225–27.

23 See *The Problems of Philosophy*, New York: Oxford University Press, 1959 (first edition 1912), ch. 5.

24 A. Plantinga, *Warrant and Proper Function*, Oxford: Oxford University Press, 1993, p. 103.

25 A. Plantinga, "*Respondeo*: *Ad* BonJour," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology*, Lanham, MD: Rowman and Littlefield, 1996, p. 341.

26 Plantinga, *Warrant and Proper Function*, p. 104.

27 Plantinga, "*Respondeo*: *Ad* BonJour," p. 341

28 Certainly he uses the term 'acquaintance' for something quite foreign; in a recent paper he asks with evident sincerity why he cannot be directly acquainted with a nearby chair. See A. Plantinga, "Direct Acquaintance?" in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman & Littlefield, 2001, p. 61. For Fumerton's reply see p. 73 ff; for BonJour's, p. 83 ff.

29 L. BonJour, "Plantinga on Knowledge and Proper Function," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman and Littlefield, 1996, p. 64.

30 L. BonJour and E. Sosa, *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, Oxford: Blackwell Publishing, 2003, p. 167.

31 A. Plantinga, "Internalism, Externalism, Defeaters and Arguments for Christian Belief," *Philosophia Christi*, series 2, 3:2, 2001, p. 390.

32 J. Cottingham *et al*. (eds and trans.) *The Philosophical Writings of Descartes*, Cambridge: Cambridge University Press, 1984, p. 14.
33 Ibid., p. 25.
34 Ibid., pp. 103–05.
35 Ibid.
36 Plantinga, *Warrant and Proper Function*, p. 109.
37 G. Frege, *The Foundations of Arithmetic*, trans. J. L. Austin, Oxford: Basil Blackwell, 1950, p. 234.
38 S. Haack, *Deviant Logic, Fuzzy Logic*: *Beyond the Formalism*, Chicago: University of Chicago Press, 1996, p. 29.
39 Putnam, "The Meaning of 'Meaning'," p. 220.
40 BonJour, *In Defense of Pure Reason*, p. 72.
41 Ibid., pp. 150–51
42 Ibid., p. 151.
43 Ibid.
44 We do not mean to deny the existence of propositions or other *abstracta* such as sets, merely to acknowledge that there are understandable questions about their existence. The existence of *abstracta* is one of the few issues on which the authors are not in entire agreement.
45 BonJour has an elaborate metaphysical theory of thoughts, arguing that they can be informed in a neo-Thomistic fashion by the forms of universals (*In Defense of Pure Reason*, pp. 180–84).
46 The frequent comparisons between concepts and experiences are not meant to indicate that concepts simply *are* experiences (perhaps experiences of a special type); we are not taking a definitive position on that question here.
47 BonJour, *In Defense of Pure Reason*, p. 41.
48 Quibbles about denotation of names aside.
49 J. Pollock, *Knowledge and Justification*, Princeton, NJ: Princeton University Press, 1974, p. 322.
50 BonJour, *In Defense of Pure Reason*, p. 151.
51 See the discussion above regarding analyticity and the theory of truth. BonJour seems to miss this point when he says (*In Defense of Pure Reason*, p. 152) that the radically skeptical possibility regarding empirical statements (e.g. that they may radically misrepresent the world of mind-independent reality) does not entail that empirical statements are not about the world. BonJour believes that this fact has a parallel in the case of *a priori* truths. However, a statement like "all tables are made entirely of oak" would be only *contingently* true in a universe without any tables and *false* in a universe where any tables were made of any other material. It can be falsified by conceivable states of affairs. By contrast, a statement like "all triangles have three sides" (given the usual meanings of the terms) would be *necessarily* true in a universe without actual triangles and also *necessarily* true regardless of the various conceivable properties of existing triangular objects.
52 T. Williamson, *Knowledge and Its Limits*, New York: Oxford University Press, 2000, p. 95.
53 Ibid., p. 95.
54 Richard Fumerton allows that one can be in pain without paying attention to that fact, and therefore without knowing that one is in pain. "Classical Foundationalism," p. 14.
55 Ibid., p. 13.
56 Williamson appears to be inconsistent on whether it is only outright belief in a falsehood that creates problems for safety or whether high confidence in a falsehood is sufficient to scuttle the possession of knowledge in epistemically similar situations. See *Knowledge and its Limits*, p. 97, where he says that it does not matter if one does not actually believe the false proposition in question and also p. 105 where he says that being "fairly confident" of a falsehood in a situation close to the case where the belief is true constitutes a problem for knowledge even when the belief is true. But contrast pp. 98–99 where he says "What incurs the charge of unreliability is believing a false proposition outright . . . ".

57 Williamson, *Knowledge and its Limits*, pp. 13, 96–105.
58 Ibid., p. 100.
59 Williamson seems aware that his appeal to reliability sounds like hand waving, and at one point he urges (*Knowledge and its Limits*, p. 102) that the reliability condition "does not depend only on brute intuition" since it also fits his conception of knowing as a mental state. In our view this consideration is worse than useless since that conception of knowing is itself hopelessly wrongheaded. But the defects in a reliability requirement are quite sufficient by themselves even without this added problem.
60 Williamson, *Knowledge and its Limits*, pp. 13, 96–97.
61 Ibid., p. 105.
62 Ibid., pp. 106–09.
63 See the references in notes 1 and 9.
64 Williamson, *Knowledge and its Limits*, p. 109.
65 Ibid., p. 14.
66 Ibid., p. 109.
67 Sellars, *Science, Perception and Reality*, pp. 131–32.

6 The Problem of Deduction

1 W. Alston, *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press, 1993, p. 116.
2 Ibid., p. 121.
3 A. Plantinga, *Warranted Christian Belief*, Oxford: Oxford University Press, 2000, p. 125.
4 A. Plantinga, "*Respondeo*: *Ad* BonJour," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology*, Lanham, MD: Rowman and Littlefield, 1996, p. 342.
5 S. Haack, *Deviant Logic, Fuzzy Logic: Beyond the Formalism*, Chicago: University of Chicago Press, 1996, p. 191.
6 M. Friedman, "Truth and Confirmation," in H. Kornblith (ed.) *Naturalizing Epistemology*, Cambridge, MA: MIT Press, 1985, p. 157. Friedman uses the term 'completeness theorem' to cover both soundness and completeness.
7 R. I. Aaron, *John Locke*, 2nd edn, Oxford: Oxford University Press, 1955, p. 223.
8 J. Locke, *An Essay Concerning Human Understanding*, vol. 2, A. C. Fraser (ed.) New York: Dover, 1959, book IV, ch. xvii, sec. 4, p. 391.
9 As Fraser notes in his running commentary. See *An Essay Concerning Human Understanding*, p. 397, n. 1.
10 Locke, *Essay*, IV, ii-iv.
11 E. Nagel and J. Newman, *Gödel's Proof*, New York: New York University Press, 1958, pp. 50–51. Note that the system under consideration is strong enough to permit the derivation of any formula from two contradictory formulas.
12 On the other end of the spectrum, we may in rare cases encounter a principle of inference so recondite that no human being is sure whether it is valid. In 1935, just four years after the publication of Kurt Gödel's famous incompleteness theorem, Gerhard Gentzen proved the consistency of the full system of elementary number theory. Gentzen's proof, however, appeals to a principle of inference known as transfinite induction that belongs neither to first order logic nor to elementary number theory itself. The embarrassing result is that no one is sure what to make of the proof. It is valid (and hence number theory is consistent) provided that transfinite induction is legitimate. But to date no one has found a convincing argument in intuitively certifiable steps that shows the validity of transfinite induction.
13 S. Haack, *Philosophy of Logics*, Cambridge: Cambridge University Press, 1978, pp. 233–37.
14 Ibid., p. 235.
15 For alternative matrices, see *Philosophy of Logics*, pp. 206–08; for Heyting's intuitionist logic, see pp. 216–20. A set of classical truth tables is an example of a finite characteristic matrix.

16 Haack, *Deviant Logics, Fuzzy Logics*, p. 9.
17 Ibid., p. 21.
18 Ibid., p. 6.
19 Ibid., p. 14.
20 Over a decade ago an earnest graduate student assured one of the authors (TM) that "in continental philosophy those three laws of logic – non-contradiction, excluded middle, and ... I can't remember the other one, but I have it written down somewhere – they don't hold."
21 Haack, *Philosophy of Logics*, p. 229.
22 This is not to deny that, if S now believes inconsistent propositions, there are analytic epistemic principles that can reveal the *correct* doxastic attitudes of a perfectly rational subject with S's foundational evidence. No doubt S is being irrational in some of his inferences or he would not believe inconsistent propositions. But valid rules of deductive inference by themselves will not clarify matters if inconsistent premises are taken as given.
23 Haack, *Philosophy of Logics*, p. 230.
24 Ibid.
25 D. Williams, "The Nature and Variety of the *A Priori*," *Analysis* 5, 1938, p. 93.

7 The Ground of Induction

1 In a 1926 lecture on "The Philosophy of Francis Bacon," reprinted in C. D. Broad, *Ethics and the History of Philosophy*, New York: Humanities Press, 1952. The comment appears on p. 143.
2 Philosophers as diverse as Karl Popper, A. J. Ayer, and Richard Fumerton have taken this position.
3 D. Hume, *An Enquiry Concerning Human Understanding*, E. Steinberg (ed.) Indianapolis: Hackett, 1977, pp. 15–25; *A Treatise of Human Nature*, 2nd ed., P. H. Nidditch (ed.) Oxford: Oxford University Press, 1978, Book I, part 3, section 11ff, pp. 124ff.
4 Hume, *Treatise*, p. 651.
5 D. C. Stove has an enlightening discussion of this point in "Hume, Probability, and Induction," *Philosophical Review* 74, 1965, pp. 160–77.
6 R. M. Sainsbury suggests that this is the sort of response we should like to give regarding induction as well; see his discussion in *Russell*, London: Routledge & Kegan Paul, 1979, p. 173.
7 On the other hand, Nelson Goodman's infamous "grue" paradox, sometimes touted as the "new riddle of induction," simply *is* the old problem of induction. For a convincing reduction of the grue paradox to the old Humean problem of induction and a rebuttal of the charge that semantic shenanigans can put "green" and "grue" on the same epistemic footing, see S. Blackburn, *Reason and Prediction*, Cambridge: Cambridge University Press, 1973, Chapter 4.
8 Critics of direct inference typically present this point as an intuition pump for skepticism about the move from sample to population in general. The classic formulation of the criticism can be found in C. D. Broad's famous 1918 paper "The Relation Between Induction and Probability," reprinted in *Induction, Probability, and Causation: Selected Papers*, Dordrecht: D. Reidel, 1968, pp. 1–52. John O. Wisdom explicitly draws the parallel between predicting the future and drawing inferences regarding the unsampled members of the population in *Foundations of Inference in Natural Science*, London: Methuen, 1952, pp. 217–18.
9 D. Williams, *The Ground of Induction*, New York: Russell & Russell, 1963, p. 39.
10 Ibid., p. 8.
11 For example, P. Horwich, *Probability and Evidence*, Cambridge: Cambridge University Press, 1982, pp. 33–34. J. L. Mackie, "A Defence of Induction," in *Logic and Knowledge*, Oxford: Oxford University Press, 1985, pp. 159–77. D. C. Stove, *The Rationality*

of Induction, Oxford: Oxford University Press, 1986, and Henry Kyburg in, e.g., *Probability and the Logic of Rational Belief*, Middletown, CT: Wesleyan University Press, 1961.

12 That general statements may be construed as limiting cases of probability statements is also stressed by R. B. Braithwaite, *Scientific Explanation*, Cambridge: Cambridge University Press, 1968, p. 152.

13 Arthur Prior gives a useful sketch of this controversy in his article "Logic, Traditional," in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan and Free Press, 1968, vol. 5, pp. 41–42.

14 For details of this calculation, see T. McGrew, "Direct Inference and the Problem of Induction," *The Monist*, 84, 2, 2001, pp. 153–78. See especially pp. 157–58.

15 This statement simplifies slightly: the upper and lower boundaries of the interval need not be equidistant from the sample frequency, as John Maynard Keynes points out in his *Treatise on Probability*, London: Macmillan, 1963, pp. 338–39.

16 A. J. Ayer, *The Central Questions of Philosophy*, New York: William Morrow and Co., 1973, p. 178.

17 J. Foster, *A. J. Ayer*, London: Routledge & Kegan Paul, 1985, p. 211. Foster brings this example up to counter a version of Bayes's Theorem, but it has more direct bearing on direct inference.

18 P. Caws, *The Philosophy of Science*, New York: D. Van Nostrand & Co., 1965, p. 265.

19 See, e.g., *Probability and the Logic of Rational Belief*, "Randomness and the Right Reference Class," *Journal of Philosophy* 74, 1977, pp. 501–21, and *Epistemology and Inference*, Minneapolis: Minnesota University Press, 1983.

20 See, e.g., G. Bhattacharyya and R. Johnson, *Statistical Concepts and Methods*, New York: Wiley, 1977, pp. 86–87. Similar definitions can be found in almost any statistics text.

21 E. Nagel, "Review of Donald Williams, *The Ground of Induction*," *Journal of Philosophy* 44, 1947, pp. 685–93. The quoted remark appears on p. 691.

22 I. Levi, "Direct Inference," *Journal of Philosophy* 74, 1977, pp. 9–10.

23 This point is raised in a slightly different form by Kyburg in "Randomness and the Right Reference Class," especially p. 515.

24 Strictly speaking, we should say "or better than random." The inference demands simply that the individual not be *less* likely to be typical than any other individual. But usually our evidence is simply that a given individual *is* a member of the set, and hence it is neither less nor more likely to be typical than any other.

25 This clarification explains one difference between this chapter and the earlier version published in *The Monist*. In that paper, premises 2* and 6* in the inductive argument above referred to S and to *a* as "random . . . with respect to" the property in question. The qualification is important, but strictly speaking it belongs at the metalevel rather than at the object level, since it is a statement about the subject's knowledge about the object or sample rather than a claim about the object or sample itself.

26 Wisdom, *Foundations of Inference in Natural Science*, p. 216.

27 This point was noted by Williams, *The Ground of Induction*, pp. 69, 149, though he unfortunately expounded it in a manner that did not sharply distinguish direct from inverse inference (see especially p. 149).

28 Under certain circumstances, symmetry *does* seem to underwrite a "maximum entropy" probability distribution. For a defense of maximum entropy methods by an objective Bayesian see Roger Rosenkrantz's books *Inference, Method and Decision: Toward a Bayesian Philosophy of Science*, Dordrecht: D. Reidel, 1977 and *Foundations and Applications of Inductive Probability*, Atascadero, CA: Ridgeview, 1981, and L. Marinoff, "A Resolution of Bertrand's Paradox," *Philosophy of Science* 61, 1994, pp. 1–24. An important new approach that preserves a large number of the intuitions that motivate objective Bayesianism is laid out in P. Walley, "Inferences from Multinomial Data: Learning About a Bag of Marbles," *Journal of the Royal Statistical Society*, Series B, 58, 1996, pp. 3–57.

29 The interval-valued approach is developed most clearly in H. Kyburg, *Science and Reason*, New York: Oxford University Press, 1990, and J. Pollock, *Nomic Probability and the Foundations of Induction*, New York: Oxford University Press, 1990.
30 H. Wang, "Notes on the Justification of Induction," *Journal of Philosophy* 44, 1947, 701–10. The quotation appears on p. 703.
31 Ibid., pp. 705–6.
32 Nagel, "Review of Donald Williams," p. 693.
33 B. Russell, *Human Knowledge: Its Scope and Limits*, New York: Simon and Schuster, 1948, p. 402.
34 This point is made forcefully in H. Kyburg, "The Justification of Induction," *Journal of Philosophy* 53, 1956, pp. 394–400.
35 M. Friedman, "Truth and Confirmation," in H. Kornblith (ed.) *Naturalizing Epistemology*, Cambridge, MA: MIT Press, 1985, pp. 147–67. See pp. 153–54.
36 C. D. Broad, "The Relation Between Induction and Probability," pp. 7–8.
37 Wisdom, *Foundations of Inference in Natural Science*, pp. 218.
38 A. J. Ayer, *Probability and Evidence*, London: Macmillan, 1973, pp. 41–42.
39 Ibid., p. 42.
40 Ibid., p. 43.
41 Wisdom, *Foundations of Inference in Natural Science*, p. 217.
42 Ibid., p. 218.
43 Herodotus, *History* vii, 10. Quoted in Keynes, p. 307. An earlier version of this chapter appeared as Timothy McGrew, "Direct Inference and the Problem of Induction," *The Monist* 84, 2001, 153–74.

Bibliography

Aaron, R. I. *John Locke*, 2nd edn, Oxford: Oxford University Press, 1955.

Adams, E. W. *A Primer of Probability Logic*, Stanford, CA: CSLI Publications, 1998.

Alston, W. *Epistemic Justification*, Ithaca, NY: Cornell University Press, 1989.

—— *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press, 1993.

—— "The 'Challenge' of Externalism," in R. Schantz (ed.) *The Externalist Challenge*, New York: de Gruyter, 2004, pp. 37–52.

Audi, R. *Belief, Justification, and Knowledge*, Belmont, CA: Wadsworth, 1988.

—— *The Structure of Justification*, Cambridge: Cambridge University Press, 1993.

—— "Contemporary Foundationalism," in L. Pojman (ed.) *The Theory of Knowledge: Classical and Contemporary Readings*, 2nd edn, Belmont, CA: Wadsworth, 1998, pp. 204–12.

Ayer, A. J. *The Central Questions of Philosophy*, New York: William Morrow and Co., 1973.

—— *Probability and Evidence*, London: Macmillan, 1973.

Baergen, R. *Contemporary Epistemology*, New York: Harcourt, Brace, 1995.

Barker, S. *Induction and Hypothesis*, Ithaca, NY: Cornell University Press, 1957.

Bergmann, M. "Externalism and Skepticism," *The Philosophical Review* 109, 2000, pp. 159–94.

—— "A Dilemma for Internalism," in T. Crisp *et al.* (eds) *Knowledge and Reality: Essays in Honor of Alvin Plantinga*, Dordrecht, Netherlands: Springer, 2006, pp. 137–77. Available HTTP: <http://web.ics.purdue.edu/~bergmann/dilemma.htm> (accessed September 30, 2005).

Bhattacharyya, G. and Johnson, R. *Statistical Concepts and Methods*, New York: Wiley, 1977.

Blackburn, S. *Reason and Prediction*, Cambridge: Cambridge University Press, 1973.

BonJour, L. *The Structure of Empirical Knowledge*, New York: Harvard University Press, 1985.

—— "Plantinga on Knowledge and Proper Function," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield, 1996, pp. 47–71.

—— *In Defense of Pure Reason*, New York: Cambridge University Press, 1998.

—— "Toward a Defense of Empirical Foundationalism," in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman & Littlefield, 2000, pp. 21–38.

—— "Replies to Pollock and Plantinga," in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman & Littlefield, 2001, pp. 79-85.

—— *Epistemology: Classic Problems and Contemporary Responses*, Lanham, MD: Rowman & Littlefield, 2002.

BonJour, L. and Sosa, E. *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, Oxford: Blackwell Publishing, 2003.

Braithwaite, R. B. *Scientific Explanation*, Cambridge: Cambridge University Press, 1968.

Broad, C. D. "The Philosophy of Francis Bacon," in *Ethics and the History of Philosophy*, New York: Humanities Press, 1952, pp. 117-42.

—— "The Relation Between Induction and Probability," in *Induction, Probability, and Causation: Selected Papers*, Dordrecht: D. Reidel, 1968, pp. 1–52.

Caws, P. *The Philosophy of Science*, New York: D. Van Nostrand & Co., 1965.

Chisholm, R. *Theory of Knowledge*, 3rd edn, Englewood Cliffs, NJ: Prentice Hall, 1989.

Cohen, L. J. *The Probable and the Provable*, Oxford: Clarendon Press, 1977.

Cohen, S. "Justification and Truth," *Philosophical Studies* 46, 1984, pp. 279–96.

Cottingham, J. *et al*. (eds and trans.) *The Philosophical Writings of Descartes*, Cambridge: Cambridge University Press, 1984.

de Morgan, A. *Formal Logic, or the Calculus of Inference, Necessary and Probable*, London: Taylor and Walton, 1847.

Dees, J. G. and Hart, J. A. "Paradox Regained: A Reply to Meyers and Stern," *The Journal of Philosophy* 71, 1974, pp. 367–72.

Dreher, J. H. "Evidence and Justified Belief," *Philosophical Studies* 25, 1974, pp. 435–39.

Fales, E. *A Defense of the Given*, Lanham, MD: Rowman & Littlefield, 1996.

Feldman, R. "An Alleged Defect in Gettier Counter-Examples," *Australasian Journal of Philosophy* 52, 1974, pp. 68–69.

—— *Epistemology*, Upper Saddle River, NJ: Prentice Hall, 2003.

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

—— *Working Without a Net: A Study of Egocentric Epistemology*, Oxford: Oxford University Press, 1993.

Foster, J. *A. J. Ayer*, London: Routledge & Kegan Paul, 1985.

Frege, G. *The Foundations of Arithmetic*, trans. J. L. Austin, Oxford: Basil Blackwell, 1950.

Friedman, M. "Truth and Confirmation," *Journal of Philosophy* 76, 1979, pp. 361–82; reprinted in Hillary Kornblith (ed.) *Naturalizing Epistemology*, Cambridge, MA: MIT Press, 1985, pp. 147–67.

Fumerton, R. "Russelling Causal Theories of Reference," in C. W. Savage and C. A. Anderson (eds) *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology*, Minneapolis: University of Minnesota Press, 1989, pp. 108–18.

—— *Metaepistemology and Skepticism*, Lanham, MD: Rowman & Littlefield, 1995.

—— "Plantinga, Warrant, and Christian Belief," *Philosophia Christi*, series 2, 3:2, 2001, pp. 341–52.

—— "Classical Foundationalism," in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman & Littlefield, 2001, pp. 3–20.

—— "Replies to Pollock and Plantinga," in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman & Littlefield, 2001, pp. 69–77.

—— "Inferential Internalism and the Presuppositions of Skeptical Arguments," in R. Schantz (ed.) *The Externalist Challenge*, New York: de Gruyter, 2004, pp. 157–67.

Gettier, E. "Is Justified True Belief Knowledge?" *Analysis* 23, 1963, pp. 121–23; reprinted in L. Pojman (ed.) *The Theory of Knowledge: Classical and Contemporary Readings*, 2nd edn, Belmont, CA: Wadsworth, 1998, pp. 142–44.

Goldman, A. I. "A Causal Theory of Knowing," *Journal of Philosophy* 64, 1967, pp. 357–72.

—— "The Internalist Conception of Justification," in P. French *et al.* (eds) *Midwest Studies in Philosophy* V, Minneapolis: University of Minnesota Press, 1980, pp. 27–52.

—— *Epistemology and Cognition*, Cambridge, MA: Harvard University Press, 1986.

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

—— *Deviant Logic, Fuzzy Logic: Beyond the Formalism*, Chicago: University of Chicago Press, 1996.

Heil, J. "Hilary Putnam," in A. P. Martinich and D. Sosa (eds) *A Companion to Analytic Philosophy*, New York: Blackwell, 2001, pp. 393–412.

Horwich, P. *Probability and Evidence*, Cambridge: Cambridge University Press, 1982.

Hume, D. *An Enquiry Concerning Human Understanding*, E. Steinberg (ed.) Indianapolis: Hackett, 1977.

—— *A Treatise of Human Nature*, 2nd edn, P. H. Nidditch (ed.) Oxford: Oxford University Press, 1978.

Keynes, J. M. *A Treatise on Probability*, London: Macmillan, 1963.

Kitcher, P. "The Naturalists Return," *The Philosophical Review* 101, 1992, pp. 53–114.

Korcz, K. A. "Recent Work on the Basing Relation," *American Philosophical Quarterly* 34, 1997, pp. 171–91.

Kornblith, H. "Beyond Foundationalism and the Coherence Theory," in H. Kornblith (ed.) *Naturalizing Epistemology*, Cambridge, MA: MIT Press, 1985, pp. 115–28.

Kyburg, H. "The Justification of Induction," *Journal of Philosophy* 53, 1956, pp. 394–400.

—— *Probability and the Logic of Rational Belief*, Middletown, CT: Wesleyan University Press, 1961.

—— "Randomness and the Right Reference Class," *Journal of Philosophy* 74, 1977, pp. 501–21.

—— *Epistemology and Inference*, Minneapolis: University of Minnesota Press, 1983.

—— *Science and Reason*, New York: Oxford University Press, 1990.

Lehrer, K. "Knowledge, Truth, and Evidence," *Analysis* 25, 1965, pp. 168–75.

Levi, I. "Direct Inference," *Journal of Philosophy* 74, 1977, pp. 5–29.

Lewis, C. I. *An Analysis of Knowledge and Valuation*, LaSalle, IL: Open Court, 1946.

Locke, J. *An Essay Concerning Human Understanding*, vol. 2, A. C. Fraser (ed.) New York: Dover, 1959.

McGrew, T. *The Foundations of Knowledge*, Lanham, MD: Littlefield Adams Books, 1995.

—— "A Defense of Classical Foundationalism," in L. Pojman (ed.) *The Theory of Knowledge: Classical and Contemporary Readings*, 2nd edn, Belmont, CA: Wadsworth, 1998, pp. 224–36.

—— "Direct Inference and the Problem of Induction," *The Monist* 84, 2001, pp. 153–78.

—— "Confirmation, Heuristics, and Explanatory Reasoning," *British Journal for the Philosophy of Science* 54, 2003, pp. 553–67.

—— "Has Plantinga Refuted the Historical Argument?" *Philosophia Christi* 6, 2004, pp. 7–26.

McGrew, T. and McGrew L. "Level Connections in Epistemology," *American Philosophical Quarterly* 34, 1997, pp. 85–94.

—— "Internalism and the Collapse of the Gettier Problem," *Journal of Philosophical Research* 23, 1998, pp. 239–56.

—— "Psychology for Armchair Philosophers," *Idealistic Studies* 28, 1998, pp. 147–57.

—— "What's Wrong With Epistemic Circularity," *Dialogue* 39, 2000, pp. 219–39.

Mackie, J. L. "A Defence of Induction," in *Logic and Knowledge*, Oxford: Oxford University Press, 1985, pp. 159–77.

Marinoff, L. "A Resolution of Bertrand's Paradox," *Philosophy of Science* 61, 1994, pp. 1–24.

Martin, M. *Atheism: A Philosophical Justification*, Philadelphia: Temple University Press, 1990.

Mates, B. *Elementary Logic*, New York: Oxford University Press, 1972.

Meyers, R. and Stern K. "Knowledge Without Paradox," *Journal of Philosophy* 70, 1973, pp. 147–60.

Miller, D. *Critical Rationalism: A Restatement and Defense*, Chicago, IL: Open Court, 1994.

Nagel, E. "Review of Donald Williams, *The Ground of Induction*," *Journal of Philosophy* 44, 1947, pp. 685–93.

Nagel, E. and Newman, J. *Gödel's Proof*, New York: New York University Press, 1958.
Nozick, R. *Philosophical Explanations*, Cambridge, MA: Belknap Press, 1981.
Plantinga, A. *Warrant: The Current Debate*, Oxford: Oxford University Press, 1993.
—— *Warrant and Proper Function*, Oxford: Oxford University Press, 1993.
—— "*Respondeo*: *Ad* BonJour," in J. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield, 1996, pp. 307–78.
—— *Warranted Christian Belief*, Oxford: Oxford University Press, 2000.
—— "Direct Acquaintance?" in M. DePaul (ed.) *Resurrecting Old-Fashioned Foundationalism*, Lanham, MD: Rowman & Littlefield, 2001, pp. 59–66.
—— "Internalism, Externalism, Defeaters and Arguments for Christian Belief," *Philosophia Christi*, series 2, 3:2, 2001, pp. 379–400.
Pollock, J. *Knowledge and Justification*, Princeton: Princeton University Press, 1974.
—— *Nomic Probability and the Foundations of Induction*, New York: Oxford University Press, 1990.
Prior, A. "Logic, Traditional," in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan and Free Press, 1968, vol. 5, pp. 34–45.
Putnam, H. "Meaning and Reference," *Journal of Philosophy* 70, 1973, pp. 699–711; reprinted in Paul K. Moser (ed.) *Reality in Focus*, Englewood Cliffs, NJ: Prentice Hall, 1990, pp. 381–88.
—— "The Meaning of 'Meaning'," in *Philosophical Papers*, vol. 2, *Mind, Language and Reality*, Cambridge: Cambridge University Press, 1975, pp. 215–71.
Reichenbach, H. *Experience and Prediction*, Chicago: University of Chicago Press, 1938.
Rosenkrantz, R. *Inference, Method and Decision: Toward a Bayesian Philosophy of Science*, Dordrecht: D. Reidel, 1977.
—— *Foundations and Applications of Inductive Probability*, Atascadero, CA: Ridgeview, 1981.
Russell, B. *Human Knowledge: Its Scope and Limits*, New York: Simon and Schuster, 1948.
—— *The Problems of Philosophy*, New York: Oxford University Press, 1959 (first edition 1912).
—— *An Inquiry Into Meaning and Truth: The William James Lectures for 1940 Delivered at Harvard University*, London: Unwin Paperbacks, 1980 (1940).
—— *The Theory of Knowledge: The 1913 Manuscript*, published as *The Collected Papers of Bertrand Russell*, vol. 7, London: George Allen and Unwin, 1983.
Russell, B. and Whitehead, A. N. *Principia Mathematica to* *96, Cambridge: Cambridge University Press, 1962.
Sainsbury, R. M. *Russell*, London: Routledge & Kegan Paul, 1979.
Sellars, W. *Science, Perception and Reality*, London: Routledge and Kegan Paul, 1963.
Shope, R. *The Analysis of Knowing*, Princeton, NJ: Princeton University Press, 1983.
Skyrms, B. "The Explication of 'X knows that *p*'," *Journal of Philosophy* 64, 1967, pp. 373–89.
Stove, D. C. "Hume, Probability, and Induction," *Philosophical Review* 74, 1965, pp. 160–77.
—— *The Rationality of Induction*, Oxford: Oxford University Press, 1986.
Swinburne, R. "The Irreducibility of Causation," *Dialectica* 51, 1997, pp. 79–92.
—— *Epistemic Justification*, Oxford: Oxford University Press, 2001.
Unger, P. "An Analysis of Factual Knowledge," *Journal of Philosophy* 65, 1968, pp. 157–70.
Walley, P. "Inferences from Multinomial Data: Learning About a Bag of Marbles," *Journal of the Royal Statistical Society*, Series B, 58, 1996, pp. 3–57.
Wang, H. "Notes on the Justification of Induction," *Journal of Philosophy* 44, 1947, pp. 701–10.

Watkins, J. *Science and Scepticism*, Princeton, NJ: Princeton University Press, 1984.
Williams, D. "The Nature and Variety of the *A Priori*," *Analysis* 5, 1938, pp. 85–94.
—— *The Ground of Induction*, New York: Russell & Russell, 1963.
Williamson, T. *Knowledge and its Limits*, New York: Oxford University Press, 2000.
Wisdom, J. O. *Foundations of Inference in Natural Science*, London: Methuen, 1952.

Index