INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600
EPISTEMOLOGICAL CONSEQUENCES OF A FACULTY PSYCHOLOGY

By

Jack Coady Lyons

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF PHILOSOPHY
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1999
As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Jack Coady Lyons entitled Epistemological Consequences of a Faculty Psychology and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of PhD in philosophy.

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Alvin I. Goldman

Date 7/12/99
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]
I am indebted to several people for helpful comments on earlier drafts of this material. Thanks to Scott Hendricks, Patrick Rysiew, and Dave Truncellito for comments on Chapter 2 and to Robert Cummins for comments on Chapter 5. Thanks to Keith Lehrer for comments on Chapters 1-3 and to John Pollock for comments on Chapters 1-5. Too many people to mention (or even to remember, for that matter) have helped me with comments, questions, and criticisms in interview rooms at the Eastern APA conference in 1998. Special thanks, however, are due to Alvin Goldman, for tirelessly reading draft after draft of all the material here (and quite a bit that never made it this far) and for providing detailed comments, suggestions, and encouragement throughout.
TABLE OF CONTENTS

TITLE ................................................................. 1
FINAL EXAMINING COMMITTEE APPROVAL FORM .......................... 2
STATEMENT BY AUTHOR .................................................. 3
ACKNOWLEDGEMENTS .................................................... 4
TABLE OF CONTENTS ..................................................... 5
LIST OF FIGURES ........................................................ 7
ABSTRACT ..................................................................... 8

CHAPTER 1: INTRODUCTION .................................................. 10
  1. The Goals of Epistemology ................................................ 14
  2. Two Senses of ‘Inferential’ ............................................... 23
  3. Itinerary .................................................................. 26

CHAPTER 2: THE PRIVILEGED EPISTEMIC STATUS OF PERCEPTUAL BELIEFS .................................................................................................................. 29
  1. Foundationalist Theories and Perceptual Beliefs ....................... 33
     1.1 Physical Object Foundationalism ...................................... 34
     1.2 Mental State Foundationalism ......................................... 37
  2. Coherentism and Perceptual Beliefs ........................................ 42
     2.1 Inductive and Explanatory Coherentism .............................. 43
     2.2 Objections ................................................................ 48
     2.3 Coherentism, Skepticism, and Circularity ............................ 56
  3. The Privileged/Nonprivileged Belief Distinction ......................... 58
     3.1 Privileged and Nonprivileged Beliefs ............................... 58
     3.2 Privileged Beliefs and Skepticism ..................................... 62
     3.3 Fundamental Ad Hoc Principles in Foundationalism .......... 66

CHAPTER 3: RELIABILISM AND THE P/N DISTINCTION .................. 74
  1. Simple Reliabilism .......................................................... 75
     1.1 Deconstructing Norman .................................................. 78
     1.2 The Problem of Privilege and Complications to Simple Reliabilism .................................................. 87
  2. Reliabilism and Nonprivileged Beliefs ..................................... 94
     2.1 Ex Ante Undermining ..................................................... 95
     2.2 Restricted Externalism .................................................. 98
  3. Two-Component Reliabilism ................................................ 100
     3.1 The P/N Distinction in Two-Component Reliabilism ............ 105
     3.2 Two-Component Reliabilism and P/N Reliabilism ............... 110

CHAPTER 4: AN ETIOLOGICAL, FACULTY-ORIENTED ACCOUNT OF THE P/N DISTINCTION ................................................................. 113
  1. The General (Family of) Account(s) ...................................... 114
     1.1 Cases of Privileged and Nonprivileged Beliefs ................. 115
     1.2 Perception vs. Reasoning; Norman vs. Nyrmoon ............... 119
     1.3 The Generic Etiological, Faculty-Oriented Account ............ 123
  2. Descriptive Epistemology and OE Faculties ........................... 124
     2.1 Conservatism and Intuitive Judgments ............................... 125
     2.2 A Commonsense Understanding of Faculties ..................... 130
     2.3 The OE Faculty Account .............................................. 133
     2.4 Objections .................................................................. 140
  3. Normative Epistemology and Opaque Systems .......................... 150
TABLE OF CONTENTS – Continued

3.1 Beyond Folk Psychology ................................................................. 151
3.2 Rendering the Epistemological Principles Less Arbitrary .................. 156
4. Reservations ..................................................................................... 161

CHAPTER 5: FACULTIES AND COGNITIVE SYSTEMS ............................. 163
1. A Conception of Cognitive Systems .................................................. 165
  1.1 Intuitive Distinctions ..................................................................... 167
  1.2 Double Dissociations and a More Precise Conception of Systems ..... 173
2. Systems and Modules ......................................................................... 187
  2.1 A Special Case of Cognitive Systems ............................................ 188
  2.2 A Special Case of OE Systems ...................................................... 195
3. Systems and Faculties ........................................................................ 197
4. The No-OE-Faculty Objection ............................................................... 204
5. Toward an Elaborative Epistemological Account ................................. 209

CHAPTER 6: CONSEQUENCES AND CONCLUSIONS .................................. 213
1. Recapitulation ...................................................................................... 213
2. General Results ................................................................................... 216
  2.1 The Structure of Epistemic Justification ......................................... 217
  2.2 Externalism ..................................................................................... 220
  2.3 Naturalized Epistemology ............................................................... 226
3. Particular Results ................................................................................ 232
  3.1 Perception and Inference ............................................................... 233
  3.2 The A Priori .................................................................................... 240
  3.3 Additional Possibilities ................................................................. 249

REFERENCES ....................................................................................... 254
LIST OF FIGURES

FIGURE 1 ................................................................. 169
FIGURE 2 ................................................................. 170
FIGURE 3 ................................................................. 171
FIGURE 4 ................................................................. 172
FIGURE 5 ................................................................. 183
FIGURE 6 ................................................................. 184
FIGURE 7 ................................................................. 185
FIGURE 8 ................................................................. 186
FIGURE 9 ................................................................. 186
ABSTRACT

Traditional epistemology has devoted much attention to the distinctions between perception and inference and between basic and non-basic beliefs. Here, I develop a different and more general distinction, between what I call "privileged" and "nonprivileged" beliefs; privileged beliefs are justifiable by means of an otherwise substandard argument, while nonprivileged beliefs require support by a generally adequate argument for their justification. I argue that even coherentists are tacitly committed to this distinction (although they may deny the existence of basic beliefs) and that one of the chief problems for simple reliabilist theories is that they imply that all beliefs are privileged. Any adequate epistemology has to count some beliefs as privileged and some as nonprivileged, and I suggest a way to modify reliabilist theories to accommodate this result.

The privileged/nonprivileged belief distinction suggests a framework theory about the structure of epistemic justification, a theory which improves on foundationalism, coherentism, and reliabilism in certain respects. Yet it raises the question of which beliefs are privileged and which are nonprivileged. I argue that whether or not a belief is privileged is determined by the etiology of that belief, and in particular, by the intrinsic nature and the etiology of the psychological faculty that produced that belief. A belief, therefore, is privileged if and only if it is the output of a certain kind of cognitive faculty, or system. Consequently, the beliefs produced by these faculties are such that it is possible to be justified in holding them even in the absence of a generally adequate argument. This does not mean that all the outputs of all such faculties are justified, for such beliefs might still require (and lack) inferential support or be subject to non-inferential requirements, like reliability. And of course, all such beliefs are potentially subject to defeat from other
justified beliefs.

The kind of cognitive faculties I have in mind includes, but is not restricted to, "modules", in Jerry Fodor's sense. The etiological, faculty-oriented view defended here argues for distinctive versions of externalism and naturalism in epistemology and holds some promise of illuminating certain traditional epistemological problems.
CHAPTER 1:
INTRODUCTION

"[I]t is not by a train of reasoning and argumentation that we come to be convinced of the existence of what we perceive; we ask no argument for the existence of the object, but that we perceive it; perception commands our belief upon its own authority, and disdains to rest its authority upon any reasoning whatsoever." — Thomas Reid (1967, p. 259)

One of the central tenets of a foundationalist theory of epistemic justification is that some beliefs can be justified without any evidential support from other beliefs. These, of course, are the basic beliefs, and the justification of all other, non-basic, beliefs is thought to derive eventually from them. This is an important claim, if true, although it is far from uncontroversial, and even those who endorse foundationalist theories of justification differ as to which beliefs are basic in this sense. An adequate defense of foundationalism would include both (i) an argument for the existence of basic beliefs and (ii) a specification of which beliefs are basic, i.e., which beliefs can be justified without evidential support from other beliefs.

My overarching goal here is to develop and defend the outlines of a view about the nature of epistemic justification which is similar in structure to, but more general than, foundationalism. The various versions of foundationalism will be subsumed as specific instances of the more general view that I have in mind. More importantly, so will every plausible version of coherentism and reliabilism. The view I defend here embodies two main theses. The first relies on a distinction between what I will call "privileged" and "nonprivileged" beliefs and consists of the claim that some beliefs are privileged and some beliefs are nonprivileged. The distinction concerns the kind or amount of inferential support that a belief requires for its justification, and the set of basic beliefs will emerge as
a subset of the set of privileged beliefs. It is in this sense that the view I defend is similar to, but more general than, foundationalism. The second thesis consists of a specification of which beliefs are privileged. These two theses are analogues of (i) and (ii) above.

A long and familiar tradition in epistemology assigns a distinctive status to perception as a source of epistemic justification. Perceptual beliefs, according to this view, occupy a privileged position among empirical beliefs in that their justification, unlike that of most empirical beliefs, does not require argument or inferential support. This claim, however, has never enjoyed anything like universal acceptance among epistemologists. While Reid (1967) and Berkeley (1948) seem to have held that our perceptual beliefs are non-inferentially justified (though for very different reasons), Descartes (1984) famously thought that our beliefs about physical objects were inferentially justified by beliefs about our own mental states and the existence of God. This debate is still with us today, as is witnessed by the ongoing debate between coherentism and foundationalism, and even within foundationalism, concerning the amount or kind of inferential support required for the justification of perceptual beliefs.

One of the goals of the present project is to shed light on this and related epistemological issues by developing a new account of the special status of perceptual beliefs, one that does not hinge on their being non-inferentially justified. What makes perceptual beliefs distinctive (even though they are almost certainly not alone in this respect) is not necessarily that they are justifiable in the absence of any inferential relations to other beliefs, but that they are held to different, relatively lax inferential standards. The claim that perceptual beliefs require no inferential support at all for their justification is simply an extreme version of this more general claim. That perceptual beliefs are subject to less rigorous inferential requirements than most other beliefs is a claim that should be
acceptable even to coherentists and to those foundationalists who think that perceptual beliefs about physical objects depend for their justification on inferential relations to beliefs about our own mental states. In fact, I will argue that these theorists are already tacitly committed to this view.

Perceptual beliefs, I will argue, are paradigmatic instances of privileged beliefs, although perceptual beliefs are not the only privileged beliefs. Perceptual beliefs, on the present account, retain a special, distinctive epistemological status, but they do not occupy a unique epistemic position. There is thus a sense in which perception is a special source of epistemic justification and a sense in which it is not, and the theory developed here will clarify both of these senses.

The fact that some other beliefs, besides perceptual beliefs, are also privileged raises the question of which beliefs are privileged and which beliefs are nonprivileged. I argue that the privileged or nonprivileged status of a belief is a matter of the etiology of the belief, even though the factors that serve to justify the belief might not be entirely etiological.¹ In particular, I maintain that a belief is privileged if and only if it is the output of a particular kind of cognitive faculty that has a particular sort of history. The kind of cognitive system I have in mind is a class that includes, but is not limited to, modules, in the sense of Fodor (1983).

This is the import of the reference to faculty psychology in the title. The term ‘faculty psychology’ will connote different things to different readers; some may see in this title hints of a Reidian epistemology, and others may see a Fodorian view concerning the

---

¹ What is required to turn a privileged belief into a justified belief is something that foundationalists, coherentists, and reliabilists will answer differently, and even privileged beliefs might require for their justification some inferential support from beliefs that have not causally contributed to the production of the belief in question.
cognitive scientific understanding of the mind. In fact, my position borrows from both Reid’s view concerning the epistemology of faculties and Fodor’s view of what faculties are. However, my own account differs substantially from the accounts of both of these authors. My main concern is to develop the position on its own strength and not to spend much time distinguishing it from the positions endorsed by Reid or by Fodor. As a consequence, there will be little exegesis of either philosopher. I do explicitly discuss the difference between the kind of cognitive faculty that I have in mind and Fodor’s conception of modularity, but this is largely in the interest of clarifying the former notion. The extent to which my position agrees with Reid’s position will barely be addressed. Several epistemologists who advocate very different views (Plantinga, 1993a; Lehrer, 1990; Coady, 1992) have claimed a significant debt to and affinity with Reid. This suggests that there is room for substantial disagreement about what Reid’s views actually are. Since I do not have the space for Reid exegesis, and since it would hardly contribute to the main points anyhow, I will simply acknowledge that the present view bears some resemblance to Reid’s, without explaining what that resemblance consists of.

Now, the notion of modularity and the more inclusive notion of what I will call “OE faculties” (cognitive faculties which are ratiocinatively opaque and have a particular etiology) are technical notions and might be thought on those grounds to be too recondite to be part of our pretheoretical conceptual repertoire. Thus, before any of the arguments for my position have been given, before the position has even been expounded in any detail, a possible objection might occur to the reader:

No epistemic theory linking justification and modularity (or any similar, technical notion) can be right, because the goal of epistemology is to elucidate our commonsense understanding of justification, our pretheoretical concept of
epistemically proper belief formation. Almost everyone has the concept of epistemic justification, although many of these people lack such scientific concepts as those of modularity and OE systems. Thus, any appeal to these latter notions is inappropriate in epistemology.

The quick answer to this objection is that although these scientific concepts are susceptible to a fair amount of precision and rigor, there are also vague and intuitive versions of them, and it is only the latter that I need to rely upon. A longer, but more satisfactory and interesting response requires a clearer specification of the goals of the present project.

1. The Goals of Epistemology

The (somewhat) recent advent of naturalized epistemology has forced many epistemologists into the business of metaepistemology, and much of this metaepistemic work has served to clarify the different epistemological projects one might have in mind in formulating an epistemological theory. Following Goldman (1992a) and Kitcher (1992), we can distinguish two main kinds of epistemological projects: the first, which Goldman calls "descriptive epistemology", attempts to capture (in some sense) our pretheoretical concepts and principles of justification and/or knowledge; the second, "normative epistemology", tries to improve on these (or perhaps entirely replace them) by offering better standards for belief-formation.²

There are at least two different approaches to descriptive epistemology that need to be distinguished. First, in "capturing" our pretheoretical concepts, one may wish to

---

² The term 'normative epistemology' is again due to Goldman. Kitcher calls this the "meliorative project", and he contrasts this with the "analytic project", which is similar to Goldman's notion of descriptive epistemology. Kitcher's target here, however, may be somewhat narrower than Goldman's, as we will see shortly, which is why I am adopting Goldman's terminology instead of Kitcher's.
provide an analysis of e.g., *justification*, in the strict sense of providing a set of necessary and sufficient conditions for someone's being justified in some belief at some time, such that the analysans is *synonymous with* the analysandum. For the purposes of this project, the analysans must retain all the vagueness of the analysandum, and it must contain no concepts which are not part of our commonsense conceptual repertoire (although meeting these two constraints is not, by itself, sufficient for synonymy). A second project aims not at synonymy but only at providing a set of metaphysically necessary and sufficient conditions for the concept in question. This is supposed to provide something that is equivalent to the relevant commonsense concept in the sense that it picks out the same class of entities in all possible worlds as does the commonsense concept. Here, however, the insistence on synonymy is relaxed; the set of conditions must retain the vagueness of the original concept (otherwise it would not pick out the same objects), but the theory need not restrict itself to concepts had by the folk (though it might still be useful to do so for other reasons). I will call these two projects the "synonymy project" and the "equivalence project", respectively. The former is what Kitcher's term 'the analytic project' suggests, although he makes fewer distinctions than I am making here, and he may have lumped both of these descriptive projects together under this label.

The normative project, again, takes as its goal the improvement of our pretheoretical epistemic concepts and principles or their replacement with better ones. While it might be that all normative epistemology has the same goal, there are at least two possible *approaches* to this project, depending on what one takes improvement to consist of. Kitcher (1992), for example, advocates a means-ends approach to normative epistemology: first we determine what our epistemic goals are, then we determine which set of methods or practices is most conducive to them.
Alternatively, one might adopt what I call a reflective equilibrium approach: we start with the epistemic principles and particular intuitions about cases that strike us as inviolable and work out an epistemology on the basis of them. Such an epistemology might be revisionist in important respects, since we might have to sacrifice some of our weaker intuitions about principles or cases in order to accommodate the stronger ones. As Goodman (1955) puts it in a slightly different context:

Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. . . . A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences (p. 67; italics in original).

It might very well be that once our commonsense principles are made sufficiently precise, we notice that some of them come into conflict with each other. Relying on the reflective equilibrium approach, we should then determine which of the conflicting principles are least central and abandon or modify them in order to ameliorate such conflict. Although I am not aware of any contemporary epistemologists who have explicitly endorsed this latter approach, as an approach to normative epistemology, this may just be because so few contemporary epistemologists are explicitly engaged in normative epistemology at all.3

There is a third epistemological project, in addition to the two discussed by Kitcher

---

3 Although I will not defend the claim here, I think that normative epistemology used to be the central project in epistemology and that these two aforementioned approaches, the means-ends approach and the reflective equilibrium approach, were exemplified by Hume and Descartes, respectively.
and Goldman. The "elaborative project", as I will call it, rests somewhere between the normative and purely descriptive projects. Here the goal is to elaborate and clarify the pretheoretical concepts, both going beyond our commonsense conceptual repertoire when necessary and cleaning up vagueness when possible. It may be, for example, that our commonsense concepts are limited in scope and that as a result, we simply do not have any clear intuitions about certain cases. According to the strictures of the elaborative project, we should not leave such indeterminacies in our theories simply because our pretheoretical concepts do not extend this far. Where our intuitions are unclear, elaborative epistemology should take a stand. The goal of the elaborative project is to improve on our pretheoretical concepts and principles by clarifying them, although, since the project is still at least partially descriptive, only by expansion, rather than revision. Some vagueness, like just what degree of justification one must possess in order to be justified, may be inevitable, but it is not a desideratum of this project to leave such vagueness intact.

The notion of clarification at issue here bears some obvious affinities to Carnap's (1950) notion of explication. However, there are some important differences that might serve to clarify my conception of elaborative epistemology. Carnapian explication involves the replacement of a prescientific concept with a similar, but more precise concept. Similarity to the prescientific concept is an important feature of a good explication, but one that is easily outweighed by other desiderata, such as fruitfulness and precision. For example, Carnap takes the scientific concept fish to be a good explication of the prescientific concept, even though the former excludes clear instances of the latter, e.g., whales and dolphins. Explication is thus more revisionist than the sort of clarification involved in the elaborative project.

Elaborative epistemology demands a heavier emphasis on similarity to the
pretheoretical concept. The elaborative project requires that any belief which our intuitions clearly categorize as justified remain classified thus, no matter how fruitful the new concept and its attendant classification scheme might be. Denying clear intuitions is reserved for normative epistemology, and even then, this is not justifiable merely on grounds of fruitfulness but is subject to the strictures of the relevant approach to normative epistemology.

Although I have followed Kitcher (1992) and Goldman (1992a) in calling these different epistemological projects, it might be more accurate to think of them as three different components of epistemological theorizing, especially if one is attracted to the reflective equilibrium approach to normative epistemology. Although my main concern here is with the descriptive project, I will also touch on the elaborative and normative projects, and I think it is worth pointing out how engaging in descriptive epistemology can provide a basis for pursuing normative epistemology. Given the above distinction between the means-ends approach and the reflective equilibrium approach, it is little surprise that Kitcher (1992) has misgivings about the relevance and interest of descriptive epistemology, for he advocates the means-ends approach. Presumably, our epistemic goals can be specified without engaging in descriptive epistemology (although see note 4), and thus an account of how to best attain them can likewise be given without engaging in descriptive epistemology.

However, the independence of descriptive and normative epistemology is not nearly as clear if we are attracted to the reflective equilibrium approach to normative epistemology.

---

4 For a different take on this issue, see Goldman (1992a). His main concern is that without engaging in descriptive epistemology and showing the similarities between the descriptive and normative theories, we will have little grounds on which to claim that the resulting normative theory is, in any important sense, an epistemic one.
The descriptive and the elaborative projects are intended to yield a specification and clarification of what epistemic principles we hold. In the process of trying to sort out various principles, we may come across some conflicting principles and/or intuitions such that something somewhere has to be abandoned. Some inconsistencies may be the price of the greater precision and clarity that elaborative epistemology requires. Elaborative epistemology, however, since it is still largely descriptive, will leave these inconsistencies intact; indeed, it should point them out.

When we turn to normative epistemology, we will delete or modify the principles that seem *ad hoc* or that lead to inconsistencies, retaining, as far as possible, those principles that have the deepest intuitive pull. Obviously, however, we cannot engage in reflective equilibrium on the basis of our epistemic principles without undertaking an examination of some sort to discover what those principles are. The difference between engaging in reflective equilibrium with and without explicitly engaging in descriptive or elaborative epistemology, then, is just the difference between conducting a sloppy or a careful examination of what our epistemic principles are. It is in this sense that I think that descriptive and elaborative epistemology can at least potentially contribute to normative epistemology.

I called the descriptive, elaborative, and normative projects “components” of epistemological theorizing, rather than “stages”, because I do not think it necessary that these components be provided strictly sequentially. Certainly, some descriptive epistemology must be done before tackling the elaborative or (reflective equilibrium) normative projects, but it is not necessary that the descriptive and elaborative accounts be fully completed before starting on the normative account. In particular, I think that the examination of scientific concepts, which will generally appear only in a normative
epistemological theory, can help us to better understand what is involved in our corresponding pretheoretical concepts and can guide the task of clarification that characterizes the elaborative project. This, in fact, will be the approach I adopt later. I will offer a sketch of a descriptive and then a normative account of the privileged/nonprivileged belief distinction in Chapter 4, and afterwards fill these in, while providing a hint at how an elaborative account might go, in Chapter 5.

The present work is aimed primarily at the descriptive project, and here I intend to address the equivalence project, rather than the synonymy project. My understanding of how a descriptive epistemological theory is to be formulated (which is heavily indebted to Goldman, 1992a) does not fit well with the traditional approach to proposing synonymous equivalences. For reasons that I will discuss in Chapter 3, the descriptive epistemological account I offer will restrict itself to pretheoretical concepts. However, these same reasons make the account an extremely unlikely candidate for expressing a synonymous equivalence. There are other reasons I pursue the equivalence project rather than the synonymy project: I have a dim view of the prospects and import of the synonymy project, and I do not think that most epistemologists are engaged in this project anyhow.

Requiring synonymy might help to make epistemology more intrinsically interesting to some people, but aside from that, there is no obvious benefit to be gained from this constraint. Elaborative and normative epistemology may require the deliverances of descriptive epistemology, but since elaborative normative epistemology do not even aim for the same extension as our folk concepts, there is no reason why they would need to start with a synonymous characterization. Thus, requiring synonymy in descriptive epistemology seems to get us no closer to realizing our larger goals of pursuing the elaborative and normative projects.
One proposed necessary condition for synonymy, which has been influential in ethics, at least, is G. E. Moore's (1903) "open question" criterion. Roughly, Moore claims that if it is an open question whether \( x \) is \( y \) (i.e., if it makes sense to wonder whether \( x \) is \( y \)), then '\( x \)' and '\( y \)' are not synonymous. But the mere fact that there is substantive disagreement in epistemology indicates that existing theories of justification and of knowledge do not express synonymies. This substantive (and reasonable) disagreement indicates that it is an open question, for any proposed analysans, whether it is the same thing as justification, and applying Moore's criterion, the "openness" of these questions indicates a failure of synonymy. The track record of epistemology to date makes it seem unlikely that anyone will produce an analysis that meets this criterion.5

This brings us to the other point, that most epistemology is not aimed at the synonymy project anyhow. This is evidenced by the fact that so many authors appeal to concepts at least as technical and arcane as the concept of modularity. It is doubtful, for instance, that my grandmother has the concept \textit{being a member of S's ultrasystem}, in the sense of Lehrer (1990).6 According to Lehrer,

A system \( M \) is a member of the ultrasystem of \( S \) at \( t \) if and only if either \( M \) is the acceptance system of \( S \) at \( t \) or results from eliminating one or more statements of the form '\( S \) accepts that \( q \)' where \( q \) is false, replacing one or more statements of the form '\( S \) accepts that \( q \)' with a statement of the form '\( S \) accepts that not-\( q \)' when \( q \) is false, or any combination of such eliminations and replacements in the acceptance

---

5 Similar considerations apply if we take interchangeability \textit{salva veritate} as a criterion of synonymy, perhaps even more obviously so.

6 That Lehrer's theory is incompatible with the constraints of the synonymy project should hardly be surprising, since he explicitly states (p. 6) that his goal is Carnapian explication, which I noted earlier was similar to that of the elaborative project.
system of $S$ at $t$ with the constraint that if $q$ logically entails $r$, which is false and also accepted, then ‘$S$ accepts that $r$’ must also be eliminated or replaced in the same way as ‘$S$ accepts that $q$’ was. (p. 194)

Neither, I suspect, does she have the concept of a self-presenting property in the sense of Chisholm (1989): “$P$ is self-presenting =Df Every property that $P$ entails includes the property of thinking” (p. 19). He goes on to explain, “One property may be said to include another if the first is necessarily such that anything that has it also has the second. And the property of being $F$ may be said to entail the property of being $G$ provided that believing something to be $F$ includes believing something to be $G$” (p. 19).

Now I do not mean to imply that ordinary people could not acquire these concepts with enough patience and interest; I am only claiming that it is doubtful that they have them now. Anyone who insists that they do have such concepts will have a hard time denying that they have the concepts of modules and OE faculties. Nor do I mean to imply that all epistemologists use such technical concepts in their theories; my point is merely that there is plenty of respectable work in epistemology that relies on concepts as least as foreign to the average person as are the cognitive scientific concepts I will rely on here. And if this is right, then it should not be taken as a serious challenge to my view that it cannot contribute to the synonymy project.

7 Nor do I mean to suggest that there is anything at all wrong with using technical concepts in epistemology; on the contrary, I think that the use of such concepts often introduces an admirable degree of precision, one to which I aspire in Chapter 5.

8 It is epistemological projects, not theories that are the issue here. When I claim that an account of justification or knowledge may appeal to concepts that ordinary agents do not have, I only mean that such a general project is legitimate. Some authors (e.g., Alston, 1997; Pollock 1986) have attacked certain theories on the grounds that the theory requires a certain type of inference for justification, and that such an inference requires having concepts or beliefs that the folk do not have. This is very different from what I am claiming. In saying that the epistemologist may legitimately appeal to certain concepts, I do
2. Two Senses of ‘Inferential’

Another clarification should be made up front. Since the distinction between privileged and nonprivileged beliefs is going to hinge on different inferential standards, I need to say what I mean by ‘inferential’. There is a psychological sense of the term and an epistemic sense. In the psychological sense, an inference is a process of reasoning that occurs over time and involves the causing or sustaining of conclusion beliefs by premise beliefs. In the epistemic sense, a belief is inferentially justified if and only if it is justified in virtue of standing in the appropriate evidential relations to other beliefs, relations of deductive, inductive, explanatory support and the like.

I intend to remain neutral with respect to the debate between foundationalism, coherentism, and reliabilism. As such, in every chapter following this one I will use ‘inference’, ‘inferentially justified’ and ‘inferential support’ only in the epistemic sense.9 (I will use both senses for the duration of this section, in order to better explain them both.) I use ‘inferential’ as opposed to, say, ‘evidential’, because according to foundationalism and reliabilism, some justification is non-inferential in that it does not rely on evidential support from other beliefs. Although all justification is presumably evidential, not all evidence is inferential, according to these theories. The relationship between inference in the psychological sense and inference in the epistemic sense is a controversial one, and again, I will remain neutral here. Still, it will clarify the distinction to explore some possible views

________________________

not, of course, mean to suggest that any use to which the epistemologist may put such concepts is therefore legitimate.

9 This is how BonJour (1985), for example, uses the term.
concerning this relation.

First of all, I think it is quite plausible to claim that not all inferences in the psychological sense count as inferences in the epistemic sense. I suspect, as do many people, that perception is a kind of unconscious inference, in the psychological sense of the term (Helmholtz, 1962). This is consistent with denying that perception is epistemically inferential. Perceptual beliefs may very well be (partially) caused by the outputs of edge detectors in low level vision, or Marr’s 2.5-D sketch (Marr, 1982; Marr & Nishihara, 1978). These are mental states which may be encoded in propositional, belief-like form (Fodor, 1975, 1985; Pylyshyn, 1973), but they are clearly not accessible to awareness. What unconscious inferences I am making at any given time is something I must rely on empirical science to tell me. And if this is true, then it is not at all clear that these inferences (again in the psychological sense) have any contribution to make to the justification of our beliefs. At the very least, any internalist should deny that they do.¹⁰

Some theorists, most notably Pollock (1986), hold that all inference in the epistemic sense is also inference in the psychological sense. This is just the claim that the only way for the belief that \( p \) to contribute justification to the belief that \( q \) is for the belief that \( p \) to (partially) cause or sustain the belief that \( q \), in such a way that the belief that \( q \) is drawn as a conclusion from the belief that \( p \). Anyone who holds this view has a causal/historical understanding of the basing relation, and anyone who defends a causal/historical theory should find this view attractive.

Not everyone, however, holds a causal/historical view concerning the basing

¹⁰ Furthermore, these inferences (in the psychological sense of the term) would generally be bad inferences, in the epistemic sense of the term. It would be inference that relied on a host of possibly unjustified assumptions (e.g., objects are almost always lit from above; objects can change their positions but almost never their size, etc.) and consistently violated the total evidence requirement (e.g., Fodor, 1983).
relation. Coherentists, in particular, will deny that all inference in the epistemic sense is also inference in the psychological sense. Any causal/historical chain of inference in the psychological sense (at least for agents with finite lifespans) must terminate in beliefs that are not caused by previous beliefs. Coherentists, however, deny that all relations of inferential support (again, evidential support from other beliefs) terminate in beliefs that are justified without such inferential support. Furthermore, not only coherentists, but some foundationalists, claim that even tacit beliefs that we have never explicitly represented, beliefs like ‘St. Thomas Aquinas never woke up naked in a zoo’, can contribute to the justification of other beliefs.

Thus, I want to use the term ‘inferential support’ in such a way that it is neutral with respect to what, if any, psychological relations are required for evidential support from other beliefs. Since I want to discuss each theory on its own terms, I want to leave open the possibility that the basing relation (and therefore inferential relations, in the epistemic sense) is not a causal relation. Hence, I will not say any more about the nature of the basing relation, since there is so little uncontroversial to say about it.\footnote{See, e.g., Alston (1988) for another explicit defense of the claim that the basing relation is a historical one. This position is implicit in any sort of process reliabilism, since what makes a belief justified, according to such a theory, is the process that produced it, i.e., how it came into being or how it came to be sustained. See also BonJour (1985) and Lehrer (1989, and especially 1990, pp. 168-172) for some arguments against a causal theory of basing.}

As a sort of corollary to the neutrality just mentioned, I will leave open the question of what counts as having evidence or having an argument for some belief. Again, there is room for a great deal of controversy here, which I wish to avoid. Some will claim that to have evidence for some belief, that evidence must be something the believer is consciously aware of at the time (Feldman, 1988). Other theorists will deny this. I will not take a stand
on this issue in what follows. And again, any further use of 'inference', 'inferential support' and the like will be intended in the epistemic sense.

3. Itinerary

Once more, there are two main theses I want to defend. The first is that some beliefs are privileged and other beliefs are nonprivileged. This occupies Chapters 2 and 3. In Chapter 2, I explain the distinction between privileged and nonprivileged beliefs and argue that all of the major internalist theories of justification are at least tacitly committed to the existence of privileged beliefs. According to all these theories, perceptual beliefs are found to occupy a privileged epistemic position, and any attempt to deny that perceptual beliefs are privileged beliefs would result in an unpalatable form of skepticism. In Chapter 3, I turn to reliabilist theories. Here, the distinction is not implicitly built into the theories, and this results in a well known, but inadequately understood, class of objections to reliabilism. These objections are groping toward an important problem with reliabilism, I suggest, and this is that it ascribes privileged status to all beliefs. A version of reliabilism that avoids these problems is sketched. Between these two chapters, the privileged/nonprivileged belief distinction is formulated and the first main thesis is defended. Chapter 2 argues that some beliefs are privileged; Chapter 3 argues that some beliefs are nonprivileged.

The second main thesis is that a belief is privileged if and only if it is the output of a certain kind of cognitive faculty. By the end of Chapter 3, the most pressing question should be the question of which beliefs are privileged and which beliefs are nonprivileged. We need an account, independent of that given in Chapters 2 and 3, of what makes some beliefs privileged and others nonprivileged. I provide the beginning of such an account in
Chapter 4.

One way to better see what this question involves is to think of a version of foundationalism (a version which I call "physical object foundationalism") which holds that perceptual beliefs and the like are basic beliefs, thus identifying the set of basic beliefs with the set of privileged beliefs. If beliefs about the physical objects I am perceiving are basic, however (i.e., if they do not require any inferential support for their justification), then it is unclear what distinguishes basic beliefs from nonbasic beliefs. Perceptual beliefs and the like are certainly not incorrigible, or self-presenting; they do not have any of the properties that a more Cartesian foundationalist tradition would have invoked to distinguish basic from nonbasic beliefs. Thus, providing an account of which beliefs are privileged and which beliefs are nonprivileged will solve -- for physical object foundationalism -- the problem of distinguishing basic from nonbasic beliefs. But the account has more general applicability than this, for I argue in Chapter 2 that all the major internalist theories are committed to the distinction between privileged and nonprivileged beliefs. Thus, the account begun in Chapter 4 will solve -- for these theories -- a problem their proponents might not have known they had.

Chapter 4 is aimed primarily at advocating a general, framework solution that I find promising: an etiological, faculty-oriented account of the privileged/nonprivileged belief distinction. My main concern there is to argue that whether or not a belief is privileged is determined by the properties of the faculty that produced the belief. However, I also want to provide a more detailed, although more tentative, account that specifies just which properties these are. Thus, in addition to the more general claim, I offer in Chapter 4 a sketch of an explicitly descriptive epistemological account of the distinction and a sketch of a normative epistemological account.
The notion of faculties plays a central role in the account of the privileged/nonprivileged belief distinction, especially in the descriptive epistemological project. (The normative epistemological account offered in Chapter 4 replaces the commonsense concept of faculties with the cognitive scientific concept of cognitive systems.) In Chapter 5, I try to develop the concept of faculties in more detail, and I do this in a somewhat roundabout way. As I mentioned above, in Section 1, I think that some pretheoretical concepts can be better understood by first getting clear on a corresponding scientific concept. In Chapter 5, I provide a detailed account of the (a?) cognitive scientific notion of cognitive systems (which is needed for the normative epistemological account), and then use this notion to get a better grasp on our pretheoretical understanding of cognitive faculties. In the course of doing this, we can see how an elaborative epistemological account might proceed.

Finally, in Chapter 6, I discuss some of the consequences I think that the two main theses have for epistemology. These two theses, when taken together, suggest some novel, though I think plausible, claims that extend beyond the scope of the problems dealt with in earlier chapters. The view that results from Chapters 2 through 5 offers new approaches to some traditional epistemological problems, and at the same time, argues for a distinctive sort of externalism (one that is consistent with most of the major internalist theories) and a sort of naturalism in epistemology that is, in some ways, more radical than what is typically urged.
CHAPTER 2:
THE PRIVILEGED EPISTEMIC STATUS OF PERCEPTUAL BELIEFS

One overarching goal of the present work is to show that the outputs of modules and similar cognitive systems have a privileged epistemic status and that this is because they share certain important properties with perceptual beliefs. Of course, this argument requires the claim that perceptual beliefs do, in fact, have a privileged epistemological status, a claim that several major epistemologies appear to deny. The purpose of this chapter is twofold. First I want to argue that all the major epistemological theories do in fact at least tacitly grant this privileged status to perceptual beliefs. What foundationalists and coherentists disagree about is not whether certain beliefs have privileged status, but on the nature of that privilege; there is more room for agreement here than one might have thought. Secondly, I will briefly point out some of the difficulties that face any epistemological theory that fails to ascribe privileged status to perceptual beliefs. Since I address externalist theories in Chapter 3, the present chapter will deal exclusively with internalist theories.

After getting some terminological matters out of the way, I will attempt to show how foundationalists and coherentists (Sections 1 and 2, respectively) ascribe a privileged epistemic status to perceptual beliefs. In Section 3 I will use the results of Sections 1 and 2 to develop in more detail the conception of this privileged class of beliefs. Although I think that there are other kinds of privileged beliefs besides perceptual beliefs (e.g., memory beliefs), for the sake of focus, the discussion will concentrate mainly on perceptual beliefs.

The present chapter relies on an intuitive understanding of what counts as a
perceptual belief; however, a bit of clarification is in order before getting started. First of all, I will take a perceptual belief to be a belief about physical objects or their properties. Secondly, I will be using the term ‘perceptual belief’ in a strict sense, according to which whether or not a particular belief is a perceptual belief is not simply a function of the proposition believed. Some beliefs that it is raining, for example, are perceptual beliefs (e.g., I look out the window and see that it is raining), but some are not (e.g., my psychic tells me it is raining and I believe it because she says so). Of course, we can sometimes tell from the content of a belief that it is not a perceptual belief (for a human); the belief that it will snow a lot next winter, for example, is simply not the sort of thing that could be a perceptual belief for a normal human. Some propositions cannot be perceptually believed (by normal humans), but no proposition, it seems, can only be perceptually believed.

There is a looser sense of ‘perceptual belief’ than the one that will be operative here, and for some purposes it is convenient to talk this way. This loose sense does classify beliefs as perceptual or not according to their contents. Some content-individuated belief types, like ‘there is something red in front of me’ have tokens that are perceptual beliefs in the strict sense, while others, like ‘nominalism is true’ do not, presumably, at least not for humans. Thus it is possible to speak in this loose sense and say that the belief that there is something red in front of me is a perceptual belief, while the belief that nominalism is true can sometimes tell from the content of a belief that it is not a perceptual belief (for a human); the belief that it will snow a lot next winter, for example, is simply not the sort of thing that could be a perceptual belief for a normal human. Some propositions cannot be perceptually believed (by normal humans), but no proposition, it seems, can only be perceptually believed.

There is a looser sense of ‘perceptual belief’ than the one that will be operative here, and for some purposes it is convenient to talk this way. This loose sense does classify beliefs as perceptual or not according to their contents. Some content-individuated belief types, like ‘there is something red in front of me’ have tokens that are perceptual beliefs in the strict sense, while others, like ‘nominalism is true’ do not, presumably, at least not for humans. Thus it is possible to speak in this loose sense and say that the belief that there is something red in front of me is a perceptual belief, while the belief that nominalism is true can sometimes tell from the content of a belief that it is not a perceptual belief (for a human); the belief that it will snow a lot next winter, for example, is simply not the sort of thing that could be a perceptual belief for a normal human. Some propositions cannot be perceptually believed (by normal humans), but no proposition, it seems, can only be perceptually believed.

1 A more precise account of perceptual belief will appear in Chapter 6. The reason for postponing this discussion is that when the time comes, I will argue that the only epistemologically relevant distinction between perceptual and non-perceptual beliefs is one that is, in part, parasitic on the privilege that accrues to perceptual beliefs. Thus, introducing such a distinction at this point would beg certain important questions. It should be noted that the distinction I will eventually turn to is between perceptual beliefs and non-perceptual beliefs, rather than between perception and non-perception. If someone perceives that $p$, then $p$; one can, however, have the perceptual belief that $p$ even though not-$p$. Thus, the sorts of analyses offered by Chisholm (1957), Dretske (1969), and Jackson (1977) are not quite relevant to the present issue.
is not a perceptual belief. When I talk here about perceptual beliefs, however, I will always have in mind the stricter sense described above.

Even given this distinction between strict and loose senses of ‘perceptual belief’ our intuitions about what counts as perceptual beliefs may not be extremely clear. Descartes (1984), for example, claimed in the 2nd Meditation that when I look out my window and claim to see people walking around on the street, all I am really seeing is cloaks and hats. Descartes classifies the belief about cloaks and hats that as a perceptual belief and the belief about people as an inferential belief. It is not obvious, to me at least, that Descartes is right about this, so there may very well be borderline cases where it is hard to determine whether a particular belief ought to be classified as a perceptual belief or not. Until the more detailed treatment is given, I will rely on what I hope are uncontroversial instances of perceptual beliefs: beliefs (held by a particular agent at a particular time) about low-level perceptible properties like shape, color, loudness, and so forth, as opposed to beliefs about, e.g., people in cloaks and hats walking around on the street under the window.

When I say that a particular belief is perceptual, I do not mean to build in any very substantive assumptions about what sorts of factors serve to justify the belief. As mentioned in Chapter 1, the nature of the basing relation, of what it is for something to be someone’s grounds or reason for some belief, is a subject of considerable controversy, and I want to remain neutral on this issue. To this end, I want to introduce the notion of a “pure perceptual belief”, by which I mean a perceptual belief (it has a certain causal history, or apparent causal history, or whatever nonsemantic property it is that makes some beliefs

---

2 Any claim about what is or is not a perceptual belief in this loose sense will have to be at least tacitly indexed to some class of cognizers (e.g., humans, or normal humans, or what have you), but it will not have to be indexed to a particular believer at a particular time. Here the property of being perceptual is attributed to the content-individuated belief type and not to any particular tokens of it.
perceptual beliefs in the strict sense), for which the agent who holds the belief has no independent evidence.

Suppose, for example, following BonJour (1985), that in order to be justified in a perceptual belief that \( p \), an agent must be justified in believing (a) that she has a cognitively spontaneous belief of kind \( K \) that \( p \), (b) that conditions \( C \) obtain, and (c) that beliefs of kind \( K \) in conditions \( C \) are (sufficiently) likely to be true. This is fully compatible with there being pure perceptual beliefs in the above sense, for suppose that \( S \) is not justified in believing that \( p \) at time \( t \); just after time \( t \), at \( t+1 \), \( S \) has the cognitively spontaneous perceptual belief that \( p \) (e.g., there's something blue in front of her) and (i) this agent is justified in believing the appropriate instances of (a) - (c) above, and (ii) the addition of this new, cognitively spontaneous perceptual belief is the only relevant change in \( S \)'s belief system between \( t \) and \( t+1 \). \( S \)'s belief that \( p \) at \( t+1 \) is a pure perceptual belief. This notion of a pure perceptual belief will be useful to keep in mind, for I intend to argue that the existence of justified pure perceptual beliefs indicates that perceptual beliefs have a privileged status, and this is tacitly recognized even by coherentists. When I talk of perceptual beliefs in what follows, I will mean pure perceptual beliefs.

Finally, my concern here is with justification, rather than knowledge, and in particular, with prima facie justification, rather than ultima facie justification. To say that a belief is prima facie justified is just to say that it has some positive epistemic support, in the sense that it is (or would be) justified, if there is (or were) no undermining or contrary evidence.

Let me begin with what I take to be a central and robust intuition: that there are different standards of justification for some of our beliefs than for others, in particular, that some of our beliefs require more substantial inferential support than others. Consider, for
example, the following two beliefs that I hold right now:

(1) my (perceptual) belief there is something red in front of me.

(2) my (nonperceptual) belief that the Andromeda Galaxy is 2.2 million light years away.

Intuitively, if I am going to be justified in holding (2) I need to have some argument, and it needs to be a good one.³ My being justified in holding (1), however, requires far less inferential support -- in fact, some versions of foundationalism would claim that such beliefs require no inferential support at all.

1. Foundationalist Theories and Perceptual Beliefs

Distinguishing two classes of beliefs according to the amount of argument or inferential support they require for their justification is not new. It has a long history in epistemology, for this is, of course, the cornerstone of a generic sort of foundationalism. Some, but not all, beliefs are justified without any argument at all, i.e.; without having any inferential connections to other beliefs. These, of course, are the basic beliefs, and all other beliefs are non-basic.

Before pursuing this any further, let me be explicit about what I take foundationalism to be, since different theorists mean different things by this term. By a "foundationalist" theory, I mean any theory according to which every justified belief is either basic or ultimately derives its justification from some set of basic justified beliefs. I will call a belief basic if and only if its prima facie justification does not require inferential relations to other beliefs. This does not mean that no inferential connections to other beliefs

³ By 'argument' I do not mean something that must be able to convince an opponent; an argument can be simply a tacit or explicit inferential relation to the belief in question.
can alter the degree of justification of a basic belief. It merely means that a basic belief has enough epistemic support (from whatever source) to count as justified even in the absence of such connections. Nor should this characterization of basic beliefs be taken to imply that all basic beliefs are justified, since it is prima facie rather than ultima facie justification that is at issue here. Thus, like BonJour (1985) and Sosa (1980), but unlike, e.g., Pollock (1986) I will classify standard externalist theories as instances of foundationalism (although they will not be discussed until the next chapter), since such theories typically claim that at least some beliefs are justified solely in virtue of some external property, in such a way that standing in the appropriate inferential relations to other beliefs is not necessary for their justification.

It is well known that foundationalists have differed with respect to just which beliefs they have taken to be basic; in fact only some versions of foundationalism view beliefs like 'there's something red in front of me' as even potentially basic. Consequently, it is hard to say anything very general about foundationalism’s view concerning basicness, expect for the fact that foundationalism makes a clear distinction between beliefs for which no argument is required and beliefs for which argument is required. There are two main kinds of (internalist) foundationalism, and I will begin with what I take to be the kind that has the most natural view concerning the status of perceptual beliefs, and which grants them the most obviously privileged epistemological status.

1.1 Physical Object Foundationalism

One kind of foundationalism, which I will call “physical object foundationalism” (henceforth POF) is the view that some beliefs about physical objects are basic. Although most externalist theories are versions of POF, it is difficult to find very many contemporary
internalist theories that take beliefs about physical objects to be basic. Probably the clearest example of such a view is that of Pollock (1986), who holds that our perceptual beliefs are justified in virtue of an evidential relation they bear to the corresponding appearance states. Thus, according to the above definition, where a basic belief is one that does not rely for its (prima facie) justification on inferential relations to other beliefs, Pollock's theory is a version of POF. Anthony Quinton (1966) defends a similar view and explicitly describes it as a version of foundationalism. Again, some basic beliefs are beliefs about physical objects, and these are justified directly, without the intervention of beliefs about how one is appeared to, by the relevant sensory or experiential state.

In Quinton's words, "not all beliefs can have other beliefs as their sufficient reasons but some must be justified by, for example, the occurrence of an experience or sense-impression" (p. 545-6). Pollock is somewhat more explicit about this. He claims that among the fundamental epistemic principles governing justification are such principles as 'x's looking red to S is a prima facie reason for S to believe that x is red' (p. 177), and 'S's seeming to remember P is a prima facie reason for S to believe P' (p. 178). Such principles are fundamental because they are neither reducible to nor derivable from more general epistemic principles: "what justifies our ordinary inference from 'x looks red to me' to 'x is red' is the simple fact that the former is a prima facie reason for the latter" (p. 44).

---

4 Some terminological differences threaten to obscure this point. Because he defines foundationalism as a doxastic view (one according to which the only things that can contribute to justification are beliefs) Pollock does not characterize his view even as a version of foundationalism, let alone POF. For Pollock, all justified beliefs are inferential, although only in an admittedly extended sense of 'inferential' -- and not the sense in use here, since some beliefs are "inferred" from non-doxastic appearance states. In categorizing Pollock as a foundationalist, I do not mean to downplay the difference between his view and that of other, more paradigmatic foundationalists.

5 Pagination here is from the version reprinted in Chisholm & Swartz (1973).
Pollock calls such reasons “noninductive nonconclusive reasons” (p. 44) because, on the one hand, they are defeasible and hence nonconclusive, and on the other hand, these reasons are not derivable from the principles governing induction (by which he means both generalization from samples and inference to the best explanation).

The reason for laying out POF in this detail is that according to this sort of view, the set of justified pure perceptual beliefs is a (strict) subset of the class of basic beliefs. POF handles the justification of pure perceptual beliefs by claiming that the physical object belief is based on and justified by the relevant appearance state in accordance with the fundamental epistemic principles governing perception. Since such beliefs are not inferentially justified by other beliefs, they are basic. Note that according to POF, basicness, like perceptual or nonperceptual status, is not determined by the content of the belief. Just because POF claims that certain propositions, like ‘There’s something red in front of me’, can be the objects of basic beliefs; this does mean that they are basic every time they are believed. My current belief that there is something red in front of me can be based solely on some argument, in which case, it is not basic, but it is not perceptual, either.

POF has a ready explanation for the above intuition: my belief that there is something red in front of me (when held on the basis of the appropriate sensory state) requires less inferential support than the belief that the Andromeda Galaxy is 2.2 million light years away, because the former belief is basic and the latter belief is not. The reason pure perceptual beliefs need less inferential support than many other beliefs is that they are basic and thus do not need any inferential support at all (where inferential support is conceived as a relation between beliefs). This, of course, does not mean that POF holds that only perceptual beliefs are basic. POF is free to claim, for example, that a priori
intuitions are basic, or that memory beliefs are basic (as does Pollock). My claim about the special status for perceptual beliefs is not the claim that perceptual beliefs are in a class by themselves; it is that they are in a class that some beliefs are not in.6

What is distinctive about POF is that it ascribes basic status to perceptual beliefs, not that it ascribes privileged status to them; other theories ascribe a privileged status to them as well. POF’s view concerning basic beliefs (that there are such things and that perceptual beliefs are among them), of course, is not shared by all internalist theories, so the next order of business is to show how perceptual beliefs can occupy this privileged position without having to be basic. For this, I need to turn to a different kind of foundationalist theory.

1.2 Mental State Foundationalism

Another, more familiar, version of foundationalism claims that the only empirical beliefs that are basic are beliefs about one’s own existence or one’s own mental states. I call this view “mental state foundationalism” (henceforth, MSF). According to this account, the only contingent beliefs that are basic are those that are about our own mental states or our own existence, and beliefs about physical objects are inferentially justified from these beliefs about our mental states, via some short inferential chain. The best-known proponents of this view are perhaps Descartes and Chisholm.7

6 POF can claim that any belief is basic, providing that POF is willing to posit an epistemic principle licensing the adoption of that kind of belief on the basis of some kind of state.

7 There is a sense in which BonJour’s (1985) doxastic presumption puts him in this camp as well. Beliefs about one’s own mental states appear on BonJour’s theory to be justified without appeal to other beliefs and hence basic. However, the coherentist element of BonJour’s theory is going to be more important here, so I will treat him as a coherentist in all ensuing discussion.
Although it may just be an oversight, Mental State Foundationalists tend to claim that the basicness of a belief is a matter of its content. Descartes held that beliefs about one’s own mental states were basic; i.e., it is the content of the belief, not why it is held, that makes it basic. Similarly, Chisholm’s (1989) notion of a self-presenting property is simply that of a property such that S’s having that property logically implies that S is thinking. Since logical implications are shared by all tokens of a belief content type, it is impossible for there to be two tokens of the same type, according to MSF, such that one token is basic and the other is not. POF, on the other hand -- or at least the plausible and extant versions of it, requires some sort of extra-doxastic relation to obtain in order for a belief to be basic. That is, it is impossible to determine from the content of the belief alone whether that belief (token) is basic or not; we must know why the belief is held before we can determine whether it is basic or not, according to POF.8

MSF claims that perceptual beliefs are inferentially supported by beliefs about our mental states. According to Chisholm, perceptual beliefs are directly inferentially supported by beliefs about our mental states. What is meant by saying that a belief is directly inferentially supported by another belief is that the inference is licensed by epistemic principles which are fundamental, in the aforementioned sense of not being

---

8 This difference between POF and MSF is actually a little more complicated than the above indicates. Chisholm (1989), for example, requires that I actually be appeared to φly in order for it to be certain for me that I’m appeared to φly. Still, this is neither an essential nor a typical feature of MSF, nor is this the sort of factor that POFists appeal to when talking about why the belief is held. Further, it is unclear what is motivating Chisholm’s hedge here. If it’s the fact that I could believe (unjustifiably) that I’m appeared to redy because my psychic told me I am, for example, then the additional requirement does not go far enough, for in this case, I could fail to be justified even if the belief happens to be true. He seems to be trying to straddle the line between the Cartesian view that basicness is a matter of the content of the belief and the Quinton/Pollock view that basicness is a matter of why the belief is held. Chisholm’s attempt to find a middle ground is quite unconvincing, and I will ignore it in what follows.
derivable from the more general principles of deduction, induction, or abduction. For many years, Chisholm (1957, 1966, 1977, 1982a, 1989) defended certain fundamental epistemic principles like that of "perceptual taking": "if \( S \) takes \( x \) to be \( f \), \( S \) has adequate evidence for . . . the proposition that \( x \) is \( f \)" (1957, p. 79).\(^9\) Chisholm insisted repeatedly that this justificatory relation is fundamental, again, in the sense of not being derivable from more general epistemic principles licensing deduction (1957, 1966, 1977), induction (1957, 1966, 1977, 1989), or abduction (1977).

For Chisholm, perceptual beliefs occupy a very different epistemic position from beliefs like 'nominalism is true', for the former are justifiable in virtue of the sort of direct inferential support described above. The latter, presumably, are not. Chisholm thinks that the justification of memory beliefs parallels that of perceptual beliefs (e.g., 1977, pp. 79-81), but that is precisely why he has to produce specific principles to deal with perceptual taking and mnemonic taking. Perception and memory are unusual. There are no general principles for taking something to be the case in the ordinary sense of 'take'. The mere fact that I take nominalism to be true, for example, does not at all provide me with adequate evidence for the belief that nominalism is true. Nor does the mere fact that I believe that I know nominalism to be true.\(^10\)

\(^9\) "Takes", for Chisholm, is a doxastic relation and not merely a matter of being appeared to in a certain way. Taking \( x \) to be \( f \) is, roughly, just believing oneself to be perceiving \( x \) to be \( f \) (1957, p. 77; Cf. 1977, p. 74).

\(^10\) One result of this account of the justification of perceptual beliefs is that, even though the basicness of a belief might be determined by the content of the belief, its privileged status is not thus determined. The proposition that there is something red in front of me is something that I can believe on the basis of a perceptual taking or on some other grounds. Only in the former case would this belief have the special status that perceptual beliefs enjoy. In this sense, MSF resembles POF: whether a particular belief token has this special epistemic status cannot always be determined merely by the content of the belief.
Nor does this view about the direct inferential support of perceptual beliefs seem to be just an idiosyncracy of Chisholm's version of mental state foundationalism. The inferential support for a perceptual belief must be either deductive, inductive, abductive, or fundamental (i.e., not derivable from any of the others). There are two historically noteworthy foundationalist attempts to found the justification of perceptual beliefs on deduction: Cartesian foundationalism and phenomenalism. According to the former, our \textit{a priori} knowledge of God's existence and the fact that he is no deceiver could be combined with appearance beliefs to yield a deductive argument for (some of) our perceptual beliefs. And according to phenomenalism, perceptual beliefs (or, more typically, statements) about physical objects were analytically implied by certain beliefs (or statements) about appearances. However, I think it is safe to say that virtually no contemporary epistemologist thinks that either of these approaches is going to be remotely satisfactory. Phenomenalism and Cartesian foundationalism are two of the most notorious failures in the history of epistemology.

Seeing no other promising deductive argument from appearances to objects, we are left with induction and abduction as the only hopeful alternatives to direct inference. How would an \textit{inductive} argument from appearances to objects go? We could argue for our perceptual beliefs by noting a past correlation between appearance beliefs and true perceptual beliefs. But what could justify our belief in the correlation? Since this sort of inductive argument would require some justified beliefs about which perceptual beliefs are true, it is clear that any such argument must be circular, and therefore not available to the foundationalist. Thus, I will postpone this view until the discussion of coherentism in Section 2. An abductive argument from appearances to objects would claim that the best explanation for my being appeared to redly or my believing that there is something red in
front of me is the fact that there is something red in front of me. Although such a position has been endorsed by foundationalists (e.g., Locke, 1975; A. H. Goldman, 1988), this view is more commonly associated with coherentism, so I will postpone this view as well until Section 2.

So, according to the standard (Chisholmian) version of mental state foundationalism, perceptual beliefs are directly inferentially supported by beliefs about our own mental states, in accordance with fundamental epistemic principles governing such inference. Since other beliefs require much more by way of argument for their justification, we see that perceptual beliefs occupy a privileged position for mental state foundationalism. This is true even though perceptual beliefs do not have the most privilege. Basic beliefs require even less inferential support (viz., none at all), but perceptual beliefs still require less inferential support (or a different kind of inferential support) than beliefs like 'nominalism is true'.

This points to an important difference between the privileged status of perceptual belief and basicness, for MSF denies that perceptual beliefs are basic but still ascribes to them a privileged status, as does POF. The fact that MSF holds that physical object beliefs are inferentially justified does not preclude attributing privileged status to them. Mental state foundationalism has a narrower conception of basic beliefs than physical object foundationalism, but it still sees perceptual beliefs as occupying a special epistemic position in that the justification of perceptual beliefs, unlike that of many other beliefs, is susceptible to a direct inference from a basic belief. This is an important finding, for it shows that it is possible to ascribe a privileged status to perceptual beliefs without ascribing basic status to them. And this shows that ascribing privileged status to perceptual beliefs is at least consistent with a coherentist theory of justification.
2. Coherentism and Perceptual Beliefs

For the sake of generality, I will define a coherence theory as any theory that is not foundationalist in the above sense. Thus, a coherence theory is any theory according to which some justified beliefs are neither basic nor ultimately derive their justification from a set of basic beliefs. Thus construed, there are probably some theories that are normally considered foundationalist or hybrid theories but would be classified as coherentist theories here. However, since I have already described foundationalist accounts of perception and memory, I will only consider in this section those theories that apply a distinctively coherentist account to the justification of perceptual beliefs.\(^{11}\)

We saw in Section 1.2 that there are two ways in which perceptual beliefs might be inferentially supported without that inferential support being either deductive or fundamental. Not surprisingly there are two main kinds of coherentist theories concerning the justification of our perceptual and memory beliefs; I will call them “inductive coherentism” and “explanatory coherentism”.

2.1 Inductive and Explanatory Coherentism

One alternative to the Chisholmian direct inference from our own mental states to the existence of external objects of perception is an inference that relies on inductive evidence. BonJour, for example, claims that perceptual beliefs are justified by tacit appeal

\(^{11}\) Conee’s (1988) view, for example, is probably a kind of coherentism on my taxonomy, even though he claims that justified beliefs must cohere with non-doxastic experiential states. Although Conee might be a coherentist in general, his view concerning perceptual beliefs seems to be similar enough to those of Quinton and Pollock to merit omission in this section. Unfortunately, it is hard to know for certain whether Conee has this sort of view in mind, since he never tells us whether experiences contribute positive epistemic support toward beliefs or merely serve as potential underiners.
to the statistical syllogism (1985, pp. 118 ff.). Consider the following example, which I will call "Defense A":

1. I have the visual belief that there is something round in front of me
2. Lighting conditions are normal, and
3. The visual beliefs I form in these conditions are almost always true.

Therefore, there is something round in front of me.12

As noted in Section 1.2, this sort of defense is not available to a foundationalist, because an argument like Defense A can only serve to justify the belief in question if the premises are themselves justified. And even though premise (i) and maybe even premise (ii) could plausibly be viewed as either basic or as deriving their justification from basic beliefs, this does not seem at all likely for premise (iii). The justification of premise (iii) is going to require justified beliefs about how often my past visual beliefs have been true, which is going to require justified beliefs about physical objects. The resulting circularity is apparent, but according to inductive coherentism, it is not a vicious circularity.

BonJour is not the only coherentist who thinks that this is how our perceptual beliefs are justified; Lehrer, in his discussions of trustworthiness (1990, 1997), seems to have something similar in mind. Both theorists, of course, think that in order for a belief to

---

12 There is an important ambiguity here that BonJour leaves without comment. BonJour's phrase is "my belief...is very likely to be true" (1985, p. 118). The best reading of "very likely" has it referring to objective probability, as this seems to be the most appropriate sort of probability. Subjective probability, in the sense of strength of belief, is simply irrelevant. 'Epistemic probability' is just a synonym for 'degree of justification', and as such, cannot support the belief that there is something round in front of me, but only the belief that I am justified in thinking that there is something round in front of me. The former does not imply the latter, since justified beliefs may be false, and, according to any view other than reliabilism, may be false more often than not. Moreover, BonJour indicates in a footnote here that the justificandum is supposed to be the belief about the physical object, not the belief about the epistemic status of a belief. Objective probability, in the sense of frequencies, propensities, or whatever, is the only sort of probability that supports the belief in question.
be *ultima facie* justified, it must also survive any undermining from other beliefs. Still, the above argument form is thought to be sufficient to confer *prima facie* justification on perceptual beliefs. Since the third premise is something we are supposed to be justified in believing on the basis of inductive evidence, I will call this sort of view "inductive coherentism".

Now, inductive coherentism, by claiming that all perceptual justification is ultimately derivable from inductive evidence in the manner just described, *appears* to be claiming that perceptual beliefs are justified by appeal to some generally adequate argument. But the general adequacy of this argument is only apparent. Consider an analogue of Defense A, namely, Defense B:

1. I have the scientific belief that the mass of the top quark is 175 GeV.
2. It's evening and I'm in my laboratory, and
3. My scientific beliefs formed in these conditions are almost always true.

Therefore, the mass of the top quark is 175 GeV.

If inductive coherentism is going to work, then something much like Defense A has to be sufficient for the *prima facie* justification of perceptual beliefs. Defense B, however, is nowhere nearly sufficient for the *prima facie* justification of the scientific belief. Even if I

---

13 Lehrer would be reluctant to put it this way, since he argues that the notion of *prima facie* justification is vacuous (1990, pp. 72-3). This is because our intuitions about which beliefs are unjustified do not distinguish between beliefs that are *prima facie* but not *ultima facie* justified and beliefs that are not even *prima facie* justified. Still, although I do not have the space to adequately argue it here, I think that there is a clear enough distinction between having a reason to believe something and not having any reason not to believe it. This is all that a notion of *prima facie* justification requires. A *prima facie* justified belief is one for which the agent has a reason to believe it, and an *ultima facie* justified belief is one that is *prima facie* justified and for which the agent does not have any reason not to believe it.
am justified in holding premise 3, the argument is inadequate. Defense B is an instance of what can be called the "Audacious Induction" (as opposed to the Pessimistic Induction); appeal to my own reliability in scientific matters is not enough.\footnote{I have omitted BonJour's qualification to cognitively spontaneous beliefs, but it does not seem as if its inclusion would alter the verdict at all.}

Similarly, even if I am justified in believing that most of my philosophical beliefs are true, this is not enough to make me even \textit{prima facie} justified in believing that nominalism is true. If I were to try to convince you that my belief in nominalism is justified by appealing to the Audacious Induction in the above manner, you would be tempted to claim that I did not have an argument for my position. Strictly speaking, however, this would be false; I would have an argument. It would just be a bad argument: most of my philosophical beliefs are true, and this is a philosophical belief. This sort of argument, however, must not be inadequate \textit{tout court} if it is to play the role prescribed for it in the justification of perceptual beliefs. This argument must be viewed as a good argument, but only for a highly constrained class of beliefs.

We see, then, that according to inductive coherentism, perceptual beliefs are held to different, lower, inferential standards than some other beliefs. This is \textit{not} meant as an objection. In fact, as I suggested in Section 2, I think that this is an important feature of any tenable (non-skeptical) epistemology. As was the case with foundationalism, memory beliefs will presumably be dealt with in the same manner as perceptual beliefs, since it is hard to see what kind of inferential support we might have for them other than appeal to our own reliability or trustworthiness in the matter.

Any coherentist account of the justification of perceptual beliefs is going to claim that perceptual beliefs are justified in virtue of having the right sort of inferential support.
from other beliefs — otherwise, perceptual beliefs would be basic. An explanatory coherence theory (e.g., Harman, 1973; Lycan, 1988) is one according to which the inferential support in question is abductive (i.e., inference to the best explanation). Does the above argument apply to these theories as well? According to explanatory coherentism, my current belief that there’s something round in front of me is justified because it is best explained by the fact that there is something round in front of me. That is, the truth of the physical object belief is justifiably inferable from the fact that I have the belief.  

But this is patently not the case with my belief that the top quark has a mass of 175 GeV. This belief is not *prima facie* justified simply because its truth best explains the fact that I believe it. If I am justified in this belief via explanatory inference, it had better be because this hypothesis best explains the collider detector’s registration of jets and positrons and their corresponding energy readings, or something of this sort. If the best explanation I could come up with for my belief about the mass of the top quark were its own truth, we would judge this belief to be unjustified, even though an analogous explanation is supposed to be adequate for perceptual beliefs. So once more, the sort of argument that is held to be sufficient for the *prima facie* justification of perceptual beliefs is seen to be nowhere nearly sufficient for the *prima facie* justification of other beliefs, and so again, we see perceptual beliefs being held to lower inferential standards.

---

15 Again, it would be natural to treat memory beliefs in the same way; presumably its having been remembered is often the best explanation for why I think I remember something happening.

16 As mentioned in Section 1.2, it is not only coherentists who claim that our perceptual beliefs are justified via explanatory inference. In fact, Locke (1975) seems to have had something like this in mind, even though the notion of abduction was not clearly formulated until much later. The difference between coherentist and foundationalist theories about the abductive support of perceptual beliefs is really just one about the justification of the explananda (i.e., whether they are foundationally justified or not), not about the explanans. Thus, the considerations marshalled here concerning the
Both inductive and explanatory coherentism see the justification of perceptual beliefs as being based partially on beliefs about our own mental states. The justification of our philosophical and scientific beliefs, on the other hand, are not justifiable in this sort of way. According to coherentism, some of our beliefs are such that the mere having of the belief provides an argument for the truth of the belief. Clearly, however, not all of our beliefs are like this: my having the belief that the mass of the top quark is 175 GeV, or that nominalism is true, does nothing to make these beliefs justified. Thus, once again, we see perceptual beliefs occupying a special epistemic position. The sorts of arguments that are sufficient for the prima facie justification of perceptual beliefs are not sufficient for the prima facie justification of belief more generally; the arguments that confer justification on perceptual beliefs are not generally adequate arguments.\footnote{In describing the justification of our perceptual beliefs, coherentists typically begin with beliefs about our beliefs about physical objects. It is possible, however, for a coherentist to claim that it is really beliefs about how we are appeared to or how things seem, instead, that enter into the inductive or abductive arguments for the truth of our perceptual and memory beliefs. The above considerations apply here as well; I do not get to infer the truth of nominalism from the fact that nominalism seems true, even though I do, presumably, get to infer that I remember that p from the fact that I seem to remember that p.}

2.2 Objections

The claim that even coherentists are tacitly committed to a distinction between privileged and nonprivileged beliefs is a surprising one, and I expect some objections to the above arguments. There are two classes of objections that I want to address here. The first sort of objection insists that coherentists really do require generally adequate arguments for all beliefs. According to this objection, the above examples only seem to show otherwise, privileged/nonprivileged distinction in explanatory coherentism should apply equally well to foundationalist theories that claim that our perceptual beliefs are abductively justified.
because the arguments on which the perceptual beliefs are based actually rely on suppressed premisses that are absent or unjustified in the other cases. The second sort of objection holds that there are other, plausible, versions of coherentism, according to which no beliefs require generally adequate arguments for their justification. I will address the first sort of objection first.

It is hard to anticipate every possible suppressed premise than an opponent might offer, but let me illustrate the general line of response by considering a few examples. First, it might be objected that the difference between Defenses A and B is that the agent is assumed to have undermining evidence in the latter case but not in the former. Justification, according to any relatively sophisticated theory, is not merely a matter of having positive evidence for some belief; it is also a matter of ruling out (in some relevant sense) evidence that threatens the positive argument. Furthermore, my claim was that perceptual belief are susceptible to relatively lax inferential standards for their \textit{prima facie} justification. The above argument would fail completely if the only difference is that the agent who relies on Defense A has a belief that is both \textit{prima facie} and \textit{ultima facie} justified, while the agent who relies on Defense B has one that is only \textit{prima facie} and not \textit{ultima facie} justified.

How might this work regarding Defenses A and B? Suppose that I have no independent evidence for thinking that the conclusion is false. Then the undermining must take the form of an argument to the effect that there is something special about this particular case. Even though beliefs of kind $K$ in conditions $C$ are usually true, some instances of conditions $C$ are also instances of condition $C^*$ (e.g., I'm in my laboratory and

\footnote{In Chapters 3 and 4, I will consider some cases which I think show that no set of missing premises will account for the distinction between beliefs that are intuitively privileged and those that are intuitively nonprivileged.}
I have not examined any data, and in these conditions, beliefs of kind $K$ are far less likely to be true. This objection can be handled simply by stipulating that the agent in question has no such evidence. The diagnosis of the cases is unchanged. Even if the agent has no beliefs at all (let alone justified beliefs) about any conditions other than $C$, Defense B remains an instance of the Audacious Induction and is not sufficient to justify the conclusion.

A common coherentist move is to claim that it is not $x$ (e.g., reliability) that makes some beliefs justified; rather it is the agent's beliefs about $x$ (reliability) that make the target belief justified. A similar approach would not be surprising here. I have argued that if inductive coherentism is true, then a mere appeal to one's own reliability or trustworthiness would have to be sufficient for the *prima facie* justification of perceptual beliefs. Since such an appeal is not sufficient for the *prima facie* justification of all beliefs, such an appeal constitutes an otherwise substandard argument. A coherentist might argue that what distinguishes my belief that there is something round in front of me from my belief that the mass of the top quark is 175 GeV is that I believe that an appeal to my own trustworthiness or reliability is sufficient for the *prima facie* justification of the former belief, but not the latter.

This response crucially invokes a *metabelief*: the belief that an appeal to one's own reliability or trustworthiness in the matter is sufficient for the *prima facie* justification of perceptual beliefs. But notice that anyone who thinks that the possession of this metabelief is necessary for the *prima facie* justification of perceptual beliefs is someone who is thereby committed to the view that the appeal to one's own reliability or trustworthiness in the matter is, in fact, insufficient for the *prima facie* justification of these beliefs, for the possession of this metabelief is also necessary. If possession of this metabelief is required
for the justification of perceptual beliefs, then this metabelief is false. It seems awkward, at best, to insist that the justification of our perceptual beliefs depends on the possession of an admittedly false belief.

Furthermore, any coherentist who would actually endorse this response will have a hard time accounting for the justification of her own perceptual beliefs. To sincerely endorse this response would be to maintain, *qua* epistemologist, that the metabelief is false (since the response insists that appeal to one's own reliability or trustworthiness in the matter is not sufficient). But at the same time, if the metabelief really is necessary, then the coherentist, *qua* epistemic agent, would have to think that the metabelief is true in order for her own perceptual beliefs to be justified.

A slightly better approach would be to concede that perceptual beliefs are, in fact, *prima facie* justifiable merely by appeal to one's own trustworthiness or reliability, but to insist that this is *because* the agent holds the metabelief. On this view, the metabelief would not actually contribute any inferential support to the perceptual belief but would only serve to distinguish those beliefs that require a generally adequate argument from those that are *prima facie* justifiable merely by appeal to one's own trustworthiness. What this would amount to is an admission of the privileged/nonprivileged distinction, coupled with an account of which beliefs are which. As such, it is perfectly consistent with the main thesis of this chapter. However, I think that this would fail to accurately distinguish privileged beliefs from nonprivileged beliefs.

It seems clear that the cognizer's beliefs about epistemology have little or no bearing on the amount of inferential support her beliefs require for their justification. Suppose that I am convinced that I do not need an argument for my belief that nominalism is true. Surely thinking does not make it so. This would make justification far too easy: if I would
only adopt the epistemic principle that all beliefs are privileged, I could become justified in all sorts of beliefs for which I lack arguments.

Suppose it only goes the other way: believing that a belief is privileged does not make it privileged, but believing that a belief is nonprivileged constitutes an underminer for Defense B. As plausible as this might initially seem, it will actually make justification too hard, for the reasons mentioned above. If inductive coherentism is right, then something like Defense A must be sufficient for the *prima facie* justification of perceptual beliefs, and it seems unlikely that normal agents have any more generally adequate arguments at their disposal. The present objection implies that agents who believe that perceptual beliefs are nonprivileged would fail to be justified in any of their perceptual beliefs. And this seems wrong. Some coherentists might believe that all beliefs are nonprivileged and that our perceptual beliefs are usually supported by generally adequate arguments. The justification of these theorists' perceptual beliefs should not be threatened by the possibility that they are wrong on both counts.

Explanatory coherentism suggests a slightly different response. One might argue that we have an additional explanation in the perceptual case, but not in the other case. That is, we have an explanation for how the truth of the perceptual belief might cause the perceptual belief, thus contributing to the strength of the explanation of the belief in terms of the fact. We do not, however, have an explanation for how the fact that the mass of the top quark is 175 GeV might cause the belief.

This objection also seems too strong. Suppose that I hear the first half of a lecture on the psychology of vision, and that the lecturer claims that there is no unique solution to the problem of inverse optics (i.e., the recovery of three dimensional objects from two dimensional retinal projections), even though the brain somehow manages to solve this
problem. I leave the lecture believing as firmly as ever that objects in the environment reliably cause my perceptual beliefs, but I am left with no idea how this might work. I am so impressed by the difficulty of the problem of inverse optics that I am no less puzzled by how objects might cause veridical perceptual beliefs than by how the mass of the top quark might cause my belief about it. Surely this puzzlement -- this lack of an explanation -- does not significantly diminish the justification of my perceptual beliefs. Certainly, it does not leave them on a par with my (unjustified) belief about the mass of the top quark.

Additionally, although my primary concern is with justification rather than knowledge, it is important to point out that such a response would be hard to reconcile with a standard coherentist theory of knowledge. If my perceptual beliefs epistemically depend in this way on beliefs about how perception works, then it looks like anyone who is mistaken about how perception works will fail to know anything on the basis of perception. One plausible necessary condition for S’s knowing that $p$ is that S’s belief that $p$ does not depend on any false beliefs for its justification (e.g., Harman, 1973; Lehrer, 1990). Aristotle had seriously false beliefs about how perception works (as might we), but surely, he at least occasionally knew that there was a table in front of him. If this is right, then his belief that there was a table in front of him must not have depended epistemically on his beliefs about how perception works.

For similar reasons, it would be implausible to maintain that our perceptual beliefs depend for their justification on our beliefs about which beliefs require what sort of inferential support. Endorsing a false epistemology should not interfere with my ability to have perceptual knowledge. But if my perceptual beliefs are epistemically dependent on my epistemological views, then, if my epistemological views are false, my perceptual beliefs will depend on false beliefs for their justification and will thus fail to count as
knowledge on a standard coherentist theory of knowledge.

The second sort of objection mentioned above is one that claims that there are versions of coherentism according to which all beliefs are, roughly, what I am calling privileged beliefs. Two kinds of coherentism that I have not yet mentioned make their appearance in epistemology textbooks: negative coherentism and holistic coherentism. Both have views about the role of reasons that seem to be inconsistent with what I am saying here. The former denies that beliefs need positive evidential support. If this is true, then there is clearly nothing special about perceptual beliefs, in that they do not require much inferential support.

Let me distinguish two kinds of negative coherentism: weak negative coherentism (a view that Harman [1986] has endorsed) denies that (most) old beliefs require positive support. This, however, is clearly consistent with claiming that all new beliefs do require some positive inferential support. Thus, if all of the above were restricted to new beliefs, then my claim about the distinctive status of perceptual beliefs would be consistent with a negative coherentism.

However, even weak negative coherentism is too strong, for it seems to be memory, rather than mere “oldness” of belief, that makes some beliefs prima facie justifiable in the absence of positive inferential support. Suppose that somewhere in the recesses of my memory is the belief that I had a black bicycle when I was ten. Without actually recalling this belief, I affirm the consequent today and merely on that basis conclude that I had a black bicycle when I was ten. We can suppose that I have no evidence against this claim, but the belief is still obviously unjustified, even though it is an old belief. It seems that all that is attractive about weak negative coherentism can be captured, more simply and with fewer awkward consequences, by some version of the
claim that memory yields privileged beliefs.

One might defend what I will call strong negative coherentism, which is weak negative coherentism plus the additional, independent claim that not even new beliefs require any positive support. I do not know of anyone who has actually held this view, and it is implausible anyhow. We have just seen a few instances of beliefs that were not even prima facie justified because they did not have enough positive support. It is hard to see how abandoning inferential requirements altogether is going to help.

Furthermore, it is unclear what might be motivating such a view. While several philosophers have argued for some principle of conservatism or other (Lycan, 1988; Harman, 1973, 1986; Quine, 1960), such a principle only applies to already held beliefs; it wouldn’t be a principle of conservatism otherwise. Neurath’s simile -- which is about repairing ships, not building them -- is perhaps more plausible than Descartes’, according to which we are like merchants who need to inspect a bushel of apples by dumping them all out and inspecting each one before replacing it. Yet this tells us nothing about the adoption of new beliefs except perhaps that it should be guided by the beliefs we already have. This does nothing to argue that any new belief is prima facie as good as any other. The very intuitions that motivate a weak negative coherentism indicate that holding on to old beliefs is quite different from adopting new ones, and thus, any negative coherentist principle concerning the justification of belief adoption needs an independent argument.

---

19 It should be noted that not every advocate of a principle of conservatism is a negative coherentist even in this weak sense. Lycan’s (1988) principle of conservatism is intended merely as a tie-breaker, the mere fact that I believe that \( p \) is not sufficient to make my belief that \( p \) even prima facie justified.

20 Even foundationalists are able to capture Neurath's intuition to some extent simply by claiming that memory beliefs are in the same category as perceptual beliefs. See Pollock (1986).
Many epistemology textbooks distinguish linear from holistic coherence theories (Audi, 1988; Baergen, 1995; Pollock, 1986). I intentionally avoided using this terminology earlier, for I think that it distracts from the central issue here. There is, in principle, at least, one kind of holistic coherence theory that threatens my claim about the special status of perceptual beliefs. According to what I will call “radically holistic” coherentism, the sole criterion for justification is coherence, where coherence is not understood as a matter of inferential support. According to this view, it is membership in a coherent set, not having certain evidential connections with other beliefs, that makes a belief justified. Such a view is suggested by some remarks that BonJour has made: “the apparent circle of justification is not in fact vicious because it is not genuinely a circle: the justification of a particular empirical belief finally depends, not on other particular beliefs as the linear conception would have it, but instead on the overall system and its coherence” (BonJour, 1985, p. 92; italics in original). If this is true, then there is nothing special about perceptual beliefs and their relatively lax inferential requirements, for no beliefs require any inferential connections in order to be justified.

Now I do not think that radically holistic coherentism is genuinely a threat to my project, partly because I do not think that there are any radically holistic coherentists. BonJour, who offers the most explicit and most radical formulation of holism, clearly does not really believe it, at least not in the sense that would undermine the present project. This is illustrated by his appeal to the Observation Argument and the neatness with which he fit into the inductive coherence category above. He obviously does think that the inferential connections that a belief has are relevant to its justification.

And it is good for him that he does not really believe in radical holistic coherentism. As Defense B illustrated, I can fail to be justified in a belief simply because I do not have an
adequate argument for it. And this intuition does not change if we stipulate that my belief about the mass of the top quark is a member of a highly coherent set, where coherence is understood as something other than inferential, evidential, connections to other beliefs. If it is possible for my belief about the mass of the top quark to be a member of a highly coherent set, despite the fact that I have no argument for this belief, then a high degree of coherence, thus understood, does not seem to be sufficient for *prima facie* justification.

### 2.3 Coherentism, Skepticism, and circularity

Before leaving the subject of coherentism, I want to consider another, more general, reason for thinking that the distinction between privileged and nonprivileged beliefs is something that is tacitly admitted by coherentism. Many discussions of coherentism introduce the view by way of the regress argument. According to the regress argument, justification must start with basic beliefs, because if it did not, then the justification for any of our beliefs would eventually either require an infinite regress of reasons or be circular, and either of these latter options would result in skepticism. What distinguishes coherentism from foundationalism is its answer to the regress argument: the coherentist denies foundationalism and skepticism by claiming that not all circularity is vicious, that some circular arguments can, in fact, confer justification.

As I will discuss in more detail below, in Section 3.2, I think that the most promising response to the skeptic is not to insist that the skeptic’s requirements have been met but to argue that the skeptic’s requirements are too strong. Coherentism, by endorsing certain kinds of circular reasoning (in certain situations), is no exception. The skeptic denies that we have any non-circular arguments for the general reliability of perception and the coherentist agrees, but denies that we *need* a non-circular argument, since circularity
does not imply viciousness.

This, however, points to another way in which coherentists tacitly accept the distinction between privileged and nonprivileged beliefs. Circular arguments, while perhaps sufficient for the justification of perceptual and memory beliefs, are not generally adequate arguments and are not taken as such. Coherentism is committed to the claim that circular arguments can serve to justify my belief that there is something round in front of me, but no plausible theory would claim that a circular argument might serve to justify my belief that God exists, or that it will snow a lot this winter, or that nominalism is true. Circular arguments can (at best) confer justification on some, but not all, beliefs.

Now of course, since coherentism eschews both foundations and infinite regresses of reasons, there is bound to be an element of circularity in every argument. However, it is important to see where this circularity would and would not be objectionable. Any argument for, say, God’s existence will presumably include a number of premise beliefs that a foundationalist would count as privileged (either perceptual beliefs like ‘there is motion’ or basic, nonempirical beliefs like ‘God is by definition that being than which a greater cannot be conceived’). Although a coherentist might plausibly claim that some of these beliefs are circularly justified, it would not be plausible to endorse any circularity that might obtain between these beliefs and the conclusion that God exists. It is in this sense that the justificatory power of circular arguments is restricted to a certain class of beliefs.

3. The Privileged/Nonprivileged Belief Distinction

Any thorough epistemological theory ought to address the following questions:

(1) How should we account for the justification of perceptual beliefs?

(2) If we have different inferential standards for perceptual beliefs than for some other beliefs, then which beliefs are subject to which set of standards?
Sections 1 and 2 canvassed the standard theories concerning the justification of perceptual belief in an effort to show that foundationalists and coherents alike posit different standards for perceptual beliefs than for some other beliefs. Although this is only implicit in coherentism, the distinction is there nonetheless. This means that all the major internalist theories answer question (1) in such a way that question (2) actually arises. (I will not begin to answer this question until Chapter 4.) Furthermore, these theories are right in positing different standards for perceptual beliefs; this is the point of the central intuition mentioned at the beginning of Section 1 and is something that will receive further defense here and in Chapter 3. There are at least two sets of standards, and it is a non-trivial issue to specify which beliefs are subject to which standards. Before even attempting this, however, we need to have a clearer understanding of the difference between the standards.

3.1 Privileged and Nonprivileged Beliefs

Since there may be other beliefs that are subject to the same sorts of standards as perceptual beliefs, we need a term for the more general class of beliefs. I will use the terms 'privileged beliefs' and 'nonprivileged beliefs' to refer, respectively, to those beliefs that are held to the special, relatively lax inferential standards for perceptual beliefs, and to the other beliefs, that are held to the other, more rigorous, inferential standards.

This terminology is only intended to convey the fact that one class of beliefs (a class that includes perceptual beliefs) is subject to less stringent inferential requirements than the other. It does not mean that perceptual beliefs and their ilk are subject to less stringent epistemic requirements for their justification, for this does not even make sense. To be justified to degree $d$ just is to have degree $d$ of epistemic support. For a belief to be justified (without the "to degree $d$" qualification) is just for it to be justified to some degree,
where that degree meets or exceeds some threshold of what counts as adequate epistemic support. To say that some belief is subject to relatively lax inferential requirements, on the other hand, is to say something about the amount or kind of inferential support such a belief must have in order to have a degree of epistemic support meeting or exceeding that threshold.

The point of distinguishing inferential standards from epistemic (i.e., justificational) standards is that it allows us to see how the privileged/nonprivileged distinction is possible. Degree of inferential support is not identical to degree of epistemic support (i.e., degree of justification). It is not as if the threshold for what counts as adequate epistemic support is different for privileged as opposed to nonprivileged beliefs; what is different is that the privileged beliefs are capable of reaching that threshold even though they are supported only by otherwise substandard arguments (or perhaps no argument at all). Nonprivileged beliefs are not.

This offers a way to characterize the distinction between privileged and nonprivileged beliefs: privileged beliefs are such that it is possible to be prima facie justified in holding them even on the basis of an otherwise substandard argument (where we take no argument at all to be the limiting case of a substandard argument). All other beliefs are nonprivileged beliefs. Since the complement of the set of otherwise substandard arguments (thus understood) is the set of generally adequate arguments, a nonprivileged belief is one whose prima facie justification requires that the agent have (in some relevant sense) a generally adequate argument.

I have not said anything about what counts as having an argument or what it is to hold a belief on the basis of some argument (although I presume that having an argument is a necessary condition for holding something on the basis of that argument). These are the
sorts of question that different theories will answer differently, depending especially on whether non-occurrent beliefs are taken to have justifying powers or not. Since, as mentioned in Chapter 1, I am trying to be ecumenical here, I will not endorse any particular reading of ‘S has argument A’ or of ‘S’s belief that p is held on the basis of A’.

Note that the above characterization of the privileged/nonprivileged (henceforth, ‘P/N’) distinction allows for the possibility of unjustified perceptual beliefs (and unjustified privileged beliefs more generally), in some importantly different ways. To say that a belief is privileged is merely to say that it is possible to be prima facie justified in holding that belief on the basis of an otherwise substandard argument. This does not mean that the possession of any such argument is sufficient for the (ultima facie) justification of privileged beliefs. First of all, an agent who holds a privileged belief on the basis of such an argument may be in possession of undermining evidence, in which case the privileged belief may not be justified (depending on whether this undermining evidence it itself undermined, etc.). Second, this otherwise substandard argument may fail to confer justification on the privileged belief either because the agent possesses the argument but does not base her belief on it, or because the premises that the argument relies on are unjustified for the agent.²¹ Even though certain otherwise substandard arguments can confer justification on some beliefs, this does not mean that such arguments cannot be

²¹ For example, one could hold a version of MSF according to which some beliefs about one’s own mental states were unjustified. Then it would be possible to be unjustified in believing that one is appeared to redly and to infer (directly) from that that there is something red nearby. This latter belief, though impeccably inferred from the appearance belief, would fail to be even prima facie justified, because the appearance belief is unjustified. It should be clear enough how an agent could fail to be justified in believing the premises of the sorts of argument that coherentists implicate in the justification of perceptual beliefs.
incorrectly applied.\textsuperscript{22}

It needs to be stressed again that the distinction between privileged and nonprivileged beliefs is not the same as the distinction between basic and non-basic beliefs. The most obvious difference is that MSF and coherentism count perceptual beliefs as privileged without classifying them as basic. More generally, privileged beliefs are ones that can be (prima facie) justified by an otherwise substandard argument, whereas basic beliefs are those that do not require any inferential support at all. Although all basic beliefs are privileged beliefs, it is possible to have privileged beliefs that are not basic, so long as it is possible for an otherwise substandard argument to confer justification. Some theorists might claim that all and only privileged beliefs are basic, but this is something that requires an argument; it is not what it means to say that a belief is privileged.

Since the P/N distinction is different from, and in a way, more general than, the basic/non-basic distinction, the former distinction makes it possible to formulate a novel hypothesis concerning the structure of epistemic justification: namely, that every belief either is privileged or ultimately derives its justification from a set of privileged beliefs. This hypothesis obviously bears some resemblance to foundationalism (it can be obtained by taking the central tenet of foundationalism and substituting 'privileged' for 'basic'), but this is not just another way of restating foundationalism, since a coherentist could cheerfully accept this hypothesis.\textsuperscript{23} In fact, a coherentist could claim that it is the truth of this hypothesis that accounts for the intuitive plausibility of foundationalist theories and that

\textsuperscript{22} As we will see in Chapter 3, reliabilism offers an additional way in which privileged beliefs might fail to be justified: they might be the result of an unreliable cognitive process.

\textsuperscript{23} This is true even if one defines coherentism, as I did above, as a denial of foundationalism.
foundationalists have simply made the mistake of conflating privilege and basicness.

3.2 Privileged Beliefs and Skepticism

The distinction between privileged and nonprivileged beliefs allows for the formulation of what I call "the P/N thesis", which is simply the claim that some beliefs are privileged and some beliefs are nonprivileged. It is not my contention here that the P/N thesis as I see it is an essential feature of either foundationalist or coherentist epistemologies. There is nothing in the definitions of foundationalism or coherentism that implies that some, but only some, beliefs are capable of being justified on the basis of an otherwise substandard argument. Instead -- and I think that this makes the P/N thesis more important than it would be if it were a trivial consequence of either of these views -- the P/N thesis as I see it is an essential feature of any plausible, non-skeptical epistemology.

There are actually two related claims here that should be separated. One is that the essential tenets of coherentism and foundationalism are consistent with the denial of the P/N thesis per se. The other is that none of the essential tenets of foundationalism or coherentism commit one to the view that the P/N distinction has the extension I have proposed here. In particular, it is possible (though implausible) to accept the P/N thesis but deny that perceptual beliefs are privileged beliefs.

It is clear enough that coherentism, per se, is neutral with respect to both the truth of the P/N thesis and the extension of the P/N distinction. Although the plausible and extant versions of coherentism treat perceptual beliefs as justifiable on the basis of otherwise substandard arguments, there is clearly nothing inconsistent about holding a coherentist view that does not. (I will discuss what is wrong with such a view shortly.) Foundationalism, on the other hand, might seem to be committed to the P/N thesis in virtue
of its account of basic beliefs, since all basic beliefs are privileged beliefs. However, the claim that there are basic beliefs does not imply that the P/N thesis is true, for P/N thesis requires, in addition, that there be some non-basic beliefs. And the P/N thesis actually requires more than just this, for there might be non-basic beliefs that are still justifiable on the basis of an otherwise substandard argument. Such beliefs would be non-basic, i.e., inferential, but still privileged. Even if the essential tenets of foundationalism were to imply that some beliefs are non-basic, this would be consistent with the claim that all beliefs are privileged. The P/N thesis insists on something stronger: that some beliefs are nonprivileged.24

More relevant to the present purposes is the fact that foundationalism, *per se*, is not committed to any particular view concerning the extension of the P/N distinction. Cartesian foundationalism, although a perfectly well-formed foundationalist theory, classifies beliefs about one’s own immediate experience as basic (and hence privileged) but sees all justified perceptual beliefs as being deduced, via some deductive argument, from them and some other premises. Since any non-circular deductive argument is going to be a generally adequate argument, Cartesian foundationalism classifies perceptual beliefs as nonprivileged beliefs.

I mentioned earlier that no one alive today thinks that Cartesian foundationalism is going to work. Now, I want to suggest that one of the main reasons for this notorious failure was that this theory failed to categorize perceptual beliefs as privileged beliefs. One of the most notorious and significant problems with Cartesian foundationalism is that it

24 Not only is foundationalism consistent with the denial of the P/N thesis, most externalist foundationalist theories defended in the literature actually *imply* the denial of the P/N thesis. This, I will argue in the next chapter, is one of the more serious problems with current versions of externalism.
made the inferential standards for the justification of perceptual beliefs too high. By requiring that perceptual beliefs be justified by appeal to a generally adequate deductive argument, this theory had the skeptical result that few if any of our perceptual beliefs are justified, since few if any of us actually have at our disposal a sound, non-circular, deductive argument for any of our perceptual beliefs.

Requiring a good deductive argument for the justification of perceptual beliefs is just one way of cashing out the claim that perceptual beliefs are nonprivileged. There are other, nondemonstrative, generally adequate arguments that one might demand of our perceptual beliefs. However, it looks as if no matter what sort of argument one might have in mind, a kind of skepticism about perception is going to result from any theory that claims that perceptual beliefs are nonprivileged. We saw in section 2 that the sorts of inductive and abductive inferences coherentists have appealed to are not generally adequate ones, for analogues of them are not nearly sufficient for the prima facie justification of some nonperceptual beliefs. Furthermore, as I argued in Section 2.3, these inferences have generally depended crucially on circularity of some kind, and circular arguments are not generally adequate. Yet these sorts of otherwise substandard arguments appear to be the only ones that normal agents actually possess. There is nothing logically inconsistent about claiming that perceptual beliefs are nonprivileged, but such a claim will leave most or all of us with a host of unjustified perceptual beliefs.

The best reason I can see for thinking that we do not possess a generally adequate argument for any of our perceptual beliefs is that the problem of skepticism with respect to the external world is still a problem. If we had a generally adequate argument for any of our perceptual beliefs, we ought to be able to use this argument in response to the skeptic.
Notoriously, no one has been able to do so. This approach to skepticism seems fruitless. The only response to skepticism that seems to hold any promise at all is the one that is embryonic in Reid (1967) and developed in more detail by contemporary epistemologists: to show, not that we have what the skeptic requires, but that the skeptic requires too much. In the case of perceptual beliefs, the only promising response to the skeptic who claims that we lack any generally adequate argument for these beliefs is to deny that perceptual beliefs need the support of a generally adequate argument. This move, in my terminology, amounts to claiming that perceptual beliefs are privileged.

Chisholm (1966,1977,1989) and Pollock (1986) have both claimed that the justification of memory beliefs parallels that of perceptual beliefs. This seems intuitively right, but it is especially plausible in light of the above discussion of skepticism. Just as we have not been able to answer the skeptic with a generally adequate argument for our perceptual beliefs, an obvious and similar, though less emphasized, problem arises concerning memory. The solution here seems to be the same as it was with perception: claim that memory beliefs are privileged, and hence that the skeptic’s request for a generally adequate argument is unreasonable.

I began with what I called a “central and robust” epistemological intuition, and we are now in a position to formulate that intuition in more precise terms: some beliefs are privileged, and some beliefs are non-privileged. This is just the P/N thesis. The foregoing discussion should buttress this intuition to the point of providing an important constraint on any adequate epistemology. Any epistemological theory that denies the existence of nonprivileged beliefs (as has been seen to some extent here but will receive extended

25 And even if some one brilliant philosopher had, it would be small comfort to the rest of us, whose perceptual beliefs would remain unjustified for us.
treatment in the following chapter) is too weak; it will falsely attribute justification to beliefs that are in fact unjustified, beliefs, e.g., that are held merely on the basis of the Audacious Induction. Conversely, as I have been arguing in this section, any theory that fails to ascribe privileged status to our perceptual beliefs will have unpalatable skeptical consequences.

3.3 Fundamental Ad Hoc Principles in Foundationalism

The difference between generally adequate (i.e., adequate in all cases) and otherwise substandard arguments is familiar enough -- we can intuitively recognize the difference. It is difficult, however, to make this characterization any more precise, absent a theory of generally adequate argument. It is still a useful characterization, for it captures in a fairly intuitive way the difference between beliefs like 'there's something red in front of me' on the one hand and 'nominalism is true' on the other, and it gives us some terminology that makes it easier to ask which other beliefs are subject to the relatively lax inferential requirements that perceptual beliefs are. We have seen that foundationalists like Pollock and Chisholm classify memory beliefs with perceptual beliefs: they are directly inferable either from a memory experience or a belief about a memory experience. It would be natural for a coherentist to make a similar claim about the relation between memory and perception, that the justification of memory beliefs relies on an inference parallel to those that support perceptual beliefs.

---

26 Note that I do not mean for general adequacy to be the same thing as formal adequacy. I doubt that there is or could be such a thing as a formal logic of nondemonstrative reasoning. Goodman (1955), I think, showed that no purely formal principles would suffice to characterize proper inductive reasoning. Still, I think that the notion of general adequacy makes at least intuitive sense, even without interpreting it as a formal property.
This raises the general question of which beliefs are privileged and which beliefs are nonprivileged. An answer to this question should involve a theory, cast in non-epistemic terms, of what makes some beliefs privileged and others nonprivileged. At the very least, we will want to do more than simply reiterate the epistemic distinctions discussed in this chapter, for it is clear that they do not provide a very clear answer to this question. Sketching such a theory is the goal of Chapters 4 and 5.

In the meantime, however, it would be helpful to try to clarify the above account -- the one given in epistemic terms -- of the P/N distinction. Since I tend to think that some kind of foundationalist theory is correct, I want to give a more precise, foundationalist, characterization of the P/N distinction. Although the account I offer is one that a coherentist could accept, foundationalism and coherentism are very different theories, and it should not surprise us if they did not agree on all the details concerning the nature of the distinction between privileged and nonprivileged beliefs.

The more precise characterization of the P/N distinction that I have in mind has already arisen in the preceding discussion of foundationalism. The principles governing perceptual belief were seen, for the foundationalist at least, to be fundamental principles in the sense that they are not derivable from other epistemic principles. This does not mean that they are not derivable from other meta-epistemic principles. This is a somewhat subtle point, so I will try approaching it from a slightly different angle. First, we should distinguish epistemic norms or rules from the factors that make some norms or rules correct.27 Epistemic norms are principles that license certain beliefs in certain conditions;

27 This distinction has been made by Goldman (1986) and by Pollock (1986). Goldman lays out the distinction in terms of J-rules vs. a criterion for J-rule rightness, while Pollock, whose focus is on different versions of externalism, distinguishes belief externalism (externalism at the level of what Goldman calls the J-rules) from norm externalism (externalism at the level of what Goldman calls the criterion for J-rule
some examples are ‘if you’re appeared to bluely, then it is permissible to believe that there is something blue’, or ‘if believing that \( p \) makes you happy, then it is permissible to believe that \( p \).

Now I take it that while the former norm is a correct one; i.e., following this norm actually produces (prima facie) justified beliefs, the latter is not correct. Thus we need some criterion for distinguishing the correct from the incorrect norms; that is, we need to specify some factor that makes some norms correct. So long as there is something other than being correct that all and only the correct norms share, then such a factor exists. Norms are (correct or incorrect) epistemic principles governing belief formation, and the criteria are metaepistemic principles which specify the factors that make some norms correct and others incorrect.

Now, when I say that some norms are fundamental, I mean that these norms cannot be derived from other norms, even though they may be subsumed by some criterion. That is, their correctness may be a consequence of some criteria for norm rightness obtaining, coupled with those norms’ having the properties that these criteria specify as being sufficient for correctness, even though they cannot be derived from any other norms that the criteria likewise specify as correct. It is in this sense of ‘fundamental’ (but possibly only in this sense) that the epistemic principles (norms) governing perceptual belief have to be fundamental, according to foundationalism.

Thus while the norm ‘if you’re appeared to bluely, then it is permissible to believe that there’s something blue nearby’ probably is not fundamental, the norm ‘if you’re appeared to perceptually, then it is permissible to believe that the corresponding perceptual rightness). This general distinction also seems to be at work, though less explicitly, in Alston (1988) and Lycan (1988). Since there is no generally accepted terminology, I will use ‘norms’, ‘rules’, and ‘epistemic principles’ interchangeably.
property is instantiated nearby’ is fundamental, according to foundationalism, because it cannot be derived from any more general epistemic norms.28 This may be true, even if there is some criterion for norm correctness that tells us what this norm and, say, the norms governing enumerative induction have in common that makes them both correct. The norm ‘if you’re appeared to bluely, then it is permissible to believe that there’s something blue nearby’ is simply a special case of the norm ‘if you’re appeared to perceptually, then it is permissible to believe that the corresponding perceptual property is instantiated nearby’, but neither is a special case of any of the norms governing induction.29

Recall the intuitive characterization of the P/N distinction: the privileged beliefs are the ones that it is possible to be justified in holding, even on the basis of otherwise substandard arguments. Foundationalism characterized the privileged beliefs as those that were either basic (POF) or directly inferable from basic beliefs (MSF) in accordance with special, fundamental epistemic principles. Both versions of foundationalism see the formation of privileged beliefs as governed by fundamental principles; the only difference is whether these principles license beliefs held on the basis of other beliefs or on the basis

28 I have cast these examples in the terms of POF, rather than MSF, simply because it takes fewer words. Analogues of these examples can be constructed for MSF simply enough.

29 Here is another way of putting all of this, in case the attempt in the text still needs clarification. There are two things that one might mean by ‘fundamental’, even though only one of these is what I will mean by ‘fundamental’. First, there is the property of not being derivable from other epistemic principles (which is what I mean by ‘fundamental’); second, there is the property of not being subsumed by any criterion of norm rightness (this is not how I will use ‘fundamental’). These two are orthogonal, so it is important to point out that saying that a norm is fundamental (in my sense) does not imply that the norm is not subsumed by any criterion of norm rightness. In particular, one might claim that there is no criterion of norm rightness; i.e., there is nothing that all and only correct epistemic principles have in common other than being correct epistemic principles. Even if this were true, it would not indicate that all norms are fundamental, in the sense of ‘fundamental’ at issue here.
of non-doxastic states. Both versions of foundationalism account for the justification of privileged beliefs by appealing to fundamental, \textit{ad hoc} principles governing a certain kind of belief formation.

When I say that these principles are \textit{ad hoc}, I do not mean this in a pejorative sense; on the contrary, I think that one of the nice features of foundationalism is that it is explicit about the \textit{ad hoc} character of the principles governing perception. Calling these principles \textit{ad hoc} is only intended to point out their special-purpose character. Any content-specific epistemic principle would be \textit{ad hoc} in this sense, for it would only apply to a certain subset of belief tokens. So, for example, if there are special epistemic principles for reasoning about other people’s mental states, these principles would be \textit{ad hoc}. Another way for principles to be \textit{ad hoc} is for them to apply only to belief tokens with a certain etiology; thus, if there are special principles governing beliefs produced by the visual system, then these principles are also \textit{ad hoc}. According to foundationalism, an argument that is adequate in some particular cases but “otherwise substandard” is typically one that is licensed in this particular instance by fundamental, \textit{ad hoc} epistemic principles. So, an initially plausible hypothesis is:

\textbf{PN*:} A belief is privileged if and only if it is licensed by epistemic principles that are both fundamental and \textit{ad hoc}; a belief is nonprivileged otherwise.

This is not quite right, but it is close, so let me explain what it means before fixing it.

Presumably, there are some epistemic principles, even if we do not know exactly what they are, governing general reasoning forms like deduction, enumerative induction, and inference to the best explanation, norms that are general in the sense of applying to any
belief token at all, regardless of the content or etiology of the belief in question.30 Due to their generality (i.e., their applicability to all belief tokens) they are not \textit{ad hoc} in the above sense. That there are such principles is perhaps most obvious in the case of deductive inference, but see, e.g., Lycan (1988) for a sketch of the general principles governing inference to the best explanation, along with a tentative argument for the claim that all non-demonstrative inference can be derived from these principles. Pollock (1990) offers a detailed theory of the general principles governing induction, and suggests (also tentatively) that inference to the best explanation can be derived from these (p. 195). These principles are fully general and govern the formation of nonprivileged beliefs.

However, not all of the norms governing the formation of nonprivileged beliefs are going to be general. For example, the norm ‘if all the swans you’ve ever seen are white, then it is permissible to infer of the next swan that it too will be white’ is \textit{ad hoc}; it only applies to a strict subset of belief tokens (viz., those tokens of the type ‘all the swans I’ve ever seen are white’). Such a principle, although \textit{ad hoc}, is not fundamental, for it is derivable from a more general principle, the one licensing the use of the statistical syllogism. This is why the privileged beliefs have to be characterized as those governed by principles that are both fundamental and \textit{ad hoc}. Of the principles governing nonprivileged beliefs, some are \textit{ad hoc} and some are fundamental, but none are both \textit{ad hoc} and fundamental.

\footnote{In claiming this, I do not mean to suggest that very much of our reasoning is the product of domain-general mechanisms. In fact, a lot of recent empirical work casts serious doubt on this. The question here concerns methods, not mechanisms. We are able to use domain-general belief forming methods on occasion (whatever the underlying mechanisms), even if we only do so for short periods of time and in highly artificial situations. Since some of the resulting beliefs are justified, there must be some epistemic principles licensing certain uses of some of these methods. This is all the above claim presupposes.}
The reason PN* does not work is that some of the principles governing the formation of privileged beliefs are going to be derived principles. Earlier I offered the example of the norm ‘if you’re appeared to bluely, then it is permissible to believe that there’s something blue nearby’, which is presumably derivable from a norm like ‘if you’re appeared to perceptually, then it is permissible to believe that the corresponding perceptual property is instantiated nearby’. If so, then only the latter norm is fundamental, even though both are ad hoc in the present sense. But this means that PN* would classify beliefs governed by the former norm as nonprivileged, rather than privileged. This problem can be fixed by offering the a somewhat more complicated account of the P/N distinction.

Let us call an epistemic principle a “P-principle” if and only if it is both fundamental and ad hoc. Then:

PN: Any belief the adoption of which is permitted by some set of (correct) P-principles is a privileged belief. Any belief the adoption of which is permitted by some epistemic principles all of which are derivable from some set of P-principles is a privileged belief. Any other belief is a nonprivileged belief.

This seems to capture the foundationalist’s characterization of the P/N distinction. This, of course, does not yet tell us which beliefs are which -- as mentioned earlier, that is the subject of Chapter 4.

On the face of it, it might appear that coherentism will have to offer a different characterization of the distinction between nonprivileged and privileged beliefs than the one offered above, since the sorts of arguments coherentists take to be sufficient for the justification of perceptual beliefs look like mere special cases of more general types of inference, like the statistical syllogism or inference to the best explanation. However, PN
is at least *consistent* with a coherentist theory of perceptual belief. The fact that the inferences that support perceptual beliefs are instances of a more general type does not imply that the principles that license such inferences are derived from the principles that license the more general type of inference. Many coherentists would probably still object to PN as a characterization of the P/N distinction, despite these comments. I do not intend to defend PN here: its use is merely expository. I offer it merely in an effort to clarify the intuitive distinction between otherwise substandard arguments and generally adequate arguments. This means, of course, that the discussion in the following chapters will have to rely on the understanding of the P/N distinction as given by the more intuitive formulation.
CHAPTER 3:
RELIABILISM AND THE PRIVILEGED/NONPRIVILEGED BELIEF DISTINCTION

The fact that the P/N distinction tacitly figures into all major internalist theories, as was argued in Chapter 2, gives some indication of the robustness of the intuitions that support the P/N thesis. All foundationalists are already committed to the existence of privileged beliefs, since all basic beliefs are privileged (although some privileged beliefs may not be basic beliefs). One of the important results from Chapter 2 is that coherentists are also committed to the existence of privileged beliefs. The plausible versions of coherentism were found to offer accounts of the justification of perceptual beliefs that relied heavily on the use of otherwise substandard arguments. The main point of Chapter 2 was to explain and illustrate the P/N distinction. But this does not yet show that the P/N thesis is true. This thesis would be false if either (i) all beliefs were privileged or (ii) all beliefs were nonprivileged. The previous chapter (Section 3.2) argued that if perceptual beliefs were not privileged, a kind of skepticism would follow, for it is hard to see what generally adequate arguments we might have for them. Yet, even if we are forced to conclude that there are privileged beliefs, the P/N thesis may be false for another reason, namely, that there are no nonprivileged beliefs.

In this chapter I want to examine the role that the P/N distinction plays in externalist theories, in particular, reliabilist theories. However, most reliabilist theories do not, even tacitly, recognize this distinction; in fact, they imply that the P/N thesis is false, as I will argue shortly, because they imply that all beliefs are privileged. Does this undercut the above argument from the ubiquity of the distinction to the robustness of the intuition that supports the P/N thesis? On the contrary; I will argue that the incompatibility of simple
versions of reliabilism with the P/N thesis constitutes an important problem for these reliabilist theories. In particular, I will try to show that one well known class of objections to reliabilist theories is best construed as attacking simple reliabilism on precisely these grounds. However, once we explicitly see these objections in terms of the privileged/nonprivileged distinction, it becomes clear how reliabilism might be amended in order to avoid this sort of difficulty.

To prevent some misunderstandings of my intentions here, I should point out that, even though I will be spending most of this chapter criticizing reliabilism, these criticisms are not intended to be decisive refutations of this general sort of view. In fact, the criticisms I want to propose here are intended to point out a way to solve some of the problems that have troubled other versions of reliabilism. Thus, the objections I will offer here are intended to eventually strengthen, rather than weaken, the case for a reliabilist theory of justification.

The following discussion is divided into three parts. In the first section, I examine relatively simple reliabilist theories, all of which err by implying that all beliefs are privileged. The standard sorts of complications to such theories are addressed, and I argue that none actually helps with this problem. In Section 2, I turn to two accounts that are explicitly aimed at the sort of problem that concerns me in the first section. In the third section I look at the more complicated account offered by Goldman (1992a). Although this theory does not quite capture the privileged/nonprivileged distinction in the right way either, I sketch what I think is a principled way to modify this view so that it does.

1. Simple Reliabilism

Let us begin with a version of reliabilism so simple that it has never (to my
knowledge) actually been endorsed in print:

(SR): A belief is *prima facie* justified if and only if it is the result of a reliable cognitive process.

(I state SR here as a version of process reliabilism, although it is easy enough to see how an analogous statement of a simple reliable indicator theory would go.)

There are many problems for such a view, one of which is that it is incomplete; without some additional clause, it tells us nothing about *ultima facie* justification.

However, since I am only concerned with *prima facie* justification, I will ignore this. The problem that concerns me here is that this view is inconsistent with the existence of nonprivileged beliefs.

SR claims that reliability is not only necessary for (*prima facie*) justification; it is sufficient. This, however, implies that no beliefs are nonprivileged, for a nonprivileged belief is simply one that it is not possible to be (even *prima facie*) justified in holding without having a generally adequate argument. Since it is always possible for there to be a reliable process which does not involve the agent's possession of a generally adequate argument, the claim that the reliability of the relevant process is sufficient for the *prima facie* justification of the belief in question amounts to the claim that all beliefs are privileged.\(^1\) This is because, recall, a privileged belief is one that is justifiable in the absence of a generally adequate argument. If it is possible to be justified in believing that \(p\), merely because the process that produced the belief that \(p\) was reliable, then it is possible

\(^1\) A more perspicuous way of putting the point is as follows. According to SR, for any proposition, a sufficient condition for being *prima facie* justified in believing that proposition is that the belief is the result of a reliable process. This implies that there is no proposition for which anything more than a reliable process is necessary, and this implies that there is no proposition for which having a generally adequate argument is necessary. Therefore, there is no proposition such that the belief in that proposition is nonprivileged.
to be justified in believing that \( p \) without having \textit{any} argument at all.\footnote{I doubt that any beliefs are such that, \textit{necessarily}, the only reliable process that could produce them is one that involves the instantiation of a generally adequate argument. But suppose that there are such beliefs. It does not follow that any such process implies the agent's \textit{possession} of that argument. A benevolent and reliable demon could infer the truth of the belief in question, and simply place the conclusion, without the argument, into the agent's mind.}

And it is even worse than this. If reliability is sufficient, then, for any belief, it is possible to be justified in holding it without having any argument at all, but this implies not only that all beliefs are \textit{privileged}, but that all beliefs are \textit{basic}, in the sense of Chapter 2, for recall that a basic belief was defined there as one that does not require any inferential support for its justification. This explains the sense in which (as mentioned in Chapter 2) (simple) reliabilist theories are versions of physical object foundationalism. If all beliefs are basic, then some beliefs about physical objects are basic.

There are two ways in which one might take issue with this consequence of simple reliabilism. First, one might -- and should -- dispute the claim that all beliefs are basic, or even the weaker claim that all beliefs are privileged. Secondly, one might deny that any beliefs about physical objects are basic (while admitting that some are privileged). This second criticism embodies a complaint against POF more generally, and consequently, I will not address it here. My concern is not with the basic/nonbasic distinction, but with the privileged/nonprivileged distinction. Thus, in what follows, I will ask the reader to suppose that some version of POF is at least plausible. That way, we can distinguish the problems with simple reliabilism from the problems with POF more generally.

Simple reliabilism's implication that all beliefs are privileged is a crippling shortcoming of the view. For it seems intuitively obvious that there are some beliefs that we cannot be \textit{even prima facie} justified in holding without having a generally adequate...
argument. Consider some of the examples from the previous chapter: my belief that nominalism is true is something for which I must have an argument if I am to be justified in holding it, even if I have no underminers for this belief. The same is true of my belief that it is going to snow a lot this winter or that the mass of the top quark is 175 GeV. Our judgment that such beliefs are unjustified when held without argument occurs before we are told *anything about* the reliability of the process that produces the belief. These intuitions indicate that having an adequate argument is (for some beliefs, at least) necessary for *prima facie* justification. And since being the result of a reliable process does not imply being based on an adequate argument, the reliability of the process must not be sufficient for the *prima facie* justification of these beliefs.

1.1 Deconstructing Norman

Although this sort of objection has not been stated in these terms before, the general idea is familiar enough and has been tacitly noted by many of the opponents of reliabilist theories (although, as I will argue, it has been misunderstood). Simple reliabilism’s incompatibility with the P/N thesis suggests the following schema for producing counterexamples to simple reliabilism: find a belief that is intuitively a nonprivileged belief but is the result of a reliable cognitive process. Simple reliabilism will count this belief as justified, and will be mistaken in so doing, whenever the agent has no generally adequate argument for the belief. For example, suppose that I have absolutely no argument for my belief that the mass of the top quark is 175 GeV, yet this belief is (unbeknownst to me) the result of a reliable process. Intuitively, this belief is not justified, even if I lack any undermining or contrary evidence.

This, I think, is the essence of such counterexamples to reliabilism as BonJour’s
(1985) Norman case and Lehrer's (1990) Truetemp example. Norman is a reliable clairvoyant, although he has no reasons for thinking so (or for thinking that he is not). One day he forms the clairvoyant belief that the president is in New York. Intuitively, Norman is not justified, even though his belief results from a reliable process. Similarly, Truetemp is an otherwise normal person who has had implanted in his head (unbeknownst to him) a Tempucomp -- a reliable device which registers the ambient temperature and produces extremely precise and accurate beliefs about the temperature. Truetemp's temperature beliefs are not justified, even though they result from a reliable process.

I want to say that the reason why we take Norman's clairvoyant belief to be unjustified is that we intuitively think it is a nonprivileged belief; that is, we intuitively think that this is one of those beliefs for which Norman needs a generally adequate argument. Similarly for Truetemp and his temperature belief. To illustrate this, consider the following case:

(1) Normina is an otherwise normal person who just happens to be extremely unreflective -- so unreflective that she has no beliefs (and a fortiori no justified beliefs) about the reliability of her own visual processes, although her visual processes are normally reliable. One day, Normina forms the visual belief (under normal lighting conditions) that there is something red in front of her.

Simple reliabilism implies that Normina's belief is prima facie justified. This, however, would serve as a terrible counterexample to simple reliabilism. Although reliabilists (e.g., Goldman, 1986, 1992a; Sosa, 1991a, 1991b) have taken the original Norman case seriously and modified their views accordingly, no one, I think, would take the present case as a serious objection to reliabilism. Intuitions may vary here, but mine is that Normina is justified. At the very least, she is less unjustified than Norman is. Note,
however, that Normina, like Norman, has no evidence for the reliability of the relevant process. Neither one of them has a generally adequate argument for his or her belief, yet one is justified (or at least nearly so) nonetheless, and the other is not. I suspect that even committed coherentists share this intuition to some extent, which is why we see cases of clairvoyance and secret neurosurgery used to argue against reliabilism instead of cases of unreflective perceivers.

I argued in the previous chapter that perceptual beliefs are privileged: they are justifiable either on the basis of an otherwise substandard argument or without any argument at all. Perhaps what distinguishes Norman’s belief from Normina’s belief is that her belief is a privileged belief and his belief is a nonprivileged belief. That would explain why, despite the doxastic similarity between the two cases, her belief is justified while his is not.

One way to support this suggestion is to notice that although the two cases are doxastically similar, Norman’s ability does not very closely resemble Normina’s ability. I have already argued that the reason perceptual beliefs are justifiable with little or no argument is that they are privileged beliefs. It is likely that what makes Norman’s belief unjustified is that it is nonprivileged, since the more dissimilar from a perceptual belief we make the clairvoyance belief, the more inclined we are to say that Norman needs an argument. In the original case, we are told only that Norman “is a completely reliable clairvoyant with respect to certain kinds of subject matter” (BonJour, 1985, p. 41). Suppose, however, that we are also told that the clairvoyant belief arises from a cognitive faculty similar to our perceptual faculties. We did not just wake up one day with fully functioning perceptual faculties, although the case BonJour offers makes it sound as if Norman’s ability has come out of nowhere.
We can consider an extreme case, in which an agent’s clairvoyance belief is even more similar to our perceptual beliefs:

(2) Nyrmoon is a member of an alien species that, several million years ago, evolved a clairvoyance module (in the sense of Fodor [1983]): a cognitive system that behaves psychologically and phenomenologically just like our perceptual systems. Nyrmoon, however, happens to be extremely unreflective and thus has no (justified) beliefs about the reliability of his clairvoyance faculty. One day, while away from home, he forms the clairvoyant belief that a friend is at his house looking for him.

Our intuitions here should be approximately as they were for the Normina case: Nyrmoon is either justified or very nearly so; in any case, he is more justified than Norman is.

However, we can alter the cases so that Norman and Nyrmoon have identical belief sets and phenomenology. Still, we tend to think that Nyrmoon is justified (or nearly so) and Norman is not.

---

3 Certain externalist theories of mental content, if true would make it difficult to specify a case where the belief sets are really identical, which is why I am not going into the details here, but I see no reason to think it cannot be done.

4 Schmitt’s (1992) Abnorman case bears a superficial resemblance to the Nyrmoon case, but there are important differences. In Schmitt’s case, Abnorman is a member of an alien species for whom clairvoyance is the norm, and Abnorman is unusual in that he has reliable visual powers as well. It is clear that Abnorman’s visual beliefs should have the same status as Norman’s clairvoyance beliefs, and Schmitt invites us to think that Abnorman’s visual beliefs would have to be justified — since they are visual beliefs, after all — and so Norman’s clairvoyance beliefs must be justified as well. I agree that Abnorman’s visual beliefs have the same status as Norman’s clairvoyance beliefs, but I have to confess that I do not have the intuition that Abnorman’s visual beliefs are justified.

The Nyrmoon case presented here is actually intended to show something quite different: that even though two agents have similar kinds of beliefs (in this case, clairvoyance beliefs), these beliefs might very well differ with respect to their epistemic status. This, I think, undermines Schmitt’s inference from the fact that Abnorman has a visual belief to the conclusion that this belief would have to be justified.
Similar considerations, of course, apply to the Truetemp case. Pit vipers, a class of snakes that includes rattlesnakes, have very sophisticated heat detection organs; suppose that something like this evolved in a humanoid species otherwise just like us. One member of this species, Vipertemp, might be doxastically and phenomenologically identical to Truetemp, but again we would want to say that Vipertemp's belief is justified even though Truetemp's is not. The most natural thing to conclude is that for some reason Norman and Truetemp need an argument while Nyrmoon and Vipertemp do not. But this is very nearly the claim that Norman's and Truetemp's beliefs are nonprivileged while Nyrmoon's and Vipertemp's beliefs are privileged.

These examples illustrate an important fact about privileged and nonprivileged beliefs. Neither Norman, Normina, nor Nyrmoon have justified beliefs in the reliability of the relevant processes. Still, some of them are justified in the beliefs that result from these processes, while the other is not. This is most clearly brought out in the contrast between Norman and Nyrmoon. Since these two are supposed to be doxastically identical, it must be our beliefs about the relevant processes (or at least our beliefs about some nondoxastic feature), not their beliefs, that cause us to ascribe privileged status to Nyrmoon's and Normina's beliefs but not to Norman's.

Now, I have intentionally formulated these cases in such a way that the beliefs in question will not be (even prima facie) justified unless they are basic, for the agents are stipulated to have no arguments whatsoever for the beliefs in question. This has two important consequences. First, it is not imperative that everyone share my intuition that Normina, Nyrmoon, and Truetemp are actually justified in their beliefs, for it is not imperative (for the present purposes) that everyone share the view that any beliefs about physical objects are ever basic. What is imperative is that the reader share the intuition that
there is an important epistemic difference between Normina and Nyrmoon, on the one 
hand, and Norman on the other hand and that Normina's and Nyrmoon's beliefs are not as 
unjustified as Norman's belief. Only readers with POF-friendly intuitions are going to 
think that Normina and Nyrmoon are justified in their beliefs (since these beliefs would 
have to be not only privileged, but basic, in order to be justified). But even those who 
deny that any physical object beliefs are basic should be able to see that Nyrmoon's belief 
has the same epistemic status as Normina's belief, and that neither of these cases would 
constitute a compelling counterexample to reliabilism.

The second important consequence of laying out the cases in this way is that the 
doxastic and phenomenological identity between Norman and Nyrmoon does not yet 
indicate that Norman's belief is nonprivileged. To show that Norman's clairvoyance 
beliefs are nonprivileged, it needs to be shown not only that agents in Norman's position 
need arguments for their beliefs, but that they need generally adequate arguments, since 
needing some inferential support is consistent with privileged status. Let us return to 
Lehrer's Truetemp example, which I think evokes stronger intuitions. Truetemp's belief is 
unjustified at least partly because he does not have any argument for it. Note, however, 
that not just any argument would help him. Suppose the following:

(3) Long before and immediately after the Tempucomp is implanted, Truetemp is 
justified in accepting that he is generally trustworthy in what he accepts.

This case actually provides a potential difficulty for coherentist theories as well, for 
Truetemp’s belief ought to be supported, to some extent, at least, by the following 
argument: 'I'm trustworthy in what I accept. I accept that it is 104 degrees outside. 
Therefore it is 104 degrees outside'. Analogous arguments are, according to a generic 
inductive coherentism (see Chapter 2), sufficient for the prima facie justification of
perceptual beliefs, even if the agent does not explicitly entertain such arguments.

Yet TrueTemp's belief in his own trustworthiness does not seem to change our intuitions about this case; he is still unjustified in his belief about the ambient temperature, even in the absence of any contrary or undermining evidence. Why is this argument not enough to *prima facie* justify TrueTemp's belief about the ambient temperature? The only apparent response is that while the argument just mentioned is adequate for perceptual beliefs, it is not generally adequate (which we already knew from Chapter 2), and TrueTemp needs a generally adequate argument for this particular belief. The most natural response here is to claim that TrueTemp's belief is a nonprivileged belief. The TrueTemp example, thus construed, would remain an objection to reliabilism without providing an argument against coherentism at the same time. And this seems intuitively right. The reason why we do not think that TrueTemp is justified is that we intuitively think that TrueTemp's belief is nonprivileged.

One might object to this by claiming that TrueTemp's belief is justified, but only *prima facie*, and that it fails to be *ultima facie* justified because of some undermining evidence that someone in TrueTemp's position could be assumed to possess. This does not strike me as very plausible, but more importantly, it is not the sort of response that an opponent of reliabilism would want to urge, for then the TrueTemp case would fail to be of

---

5 The reason I call this a potential "difficulty", rather than a counterexample, is that, unlike simple reliabilism, which implies that all beliefs are privileged, coherentism *per se* does not imply anything about which beliefs are privileged and which beliefs are not (Cf. Chapter 2, Section 3.2).

6 Analogous arguments apply to the Norman case. Perhaps our intuitions here (that the agent is still unjustified despite the metajustification) are not as decisive as they are in the TrueTemp case, but if so, I would maintain that this is merely because TrueTemp is more obviously unjustified than Norman, in the original examples.
any use as a counterexample to reliabilism. Any reasonably sophisticated statement of reliabilism makes reliability only a sufficient condition for \textit{prima facie} justification and includes a no-undermining clause to get from that to \textit{ultima facie} justification. What distinguishes Norman from BonJour's other cases is that he has no evidence for or against the claim that he is a reliable clairvoyant; the Truetemp case must be understood in the same way if it is going to provide a counterexample to the relevant sorts of reliabilist theories. Anyone who wants to claim that Truetemp's belief is \textit{prima facie} but not \textit{ultima facie} justified thus loses the Truetemp case as an argument against existing versions of reliabilism. Furthermore, we can simply stipulate that the agent has no such undermining evidence, and this does not seem to change our intuitions about the justifiedness of Truetemp's belief.

Sosa (1991b) calls the Norman and Truetemp cases instances of the "meta-incoherence problem" (p. 132). This problem, he claims, shows that one necessary condition on justification is that a belief be "internally justified" (p. 138). Similarly, BonJour takes the Norman case to show that "subjective rationality is essential to justification" (1985, p. 51). Now although neither Sosa nor BonJour is very explicit about what "internal justification" or "subjective rationality" mean, what they seem to have in mind is some argument for thinking that the belief is the product of some reliable source. Lehrer (1990) makes a similar claim after discussing Truetemp, although his concern in that discussion is primarily with knowledge rather than justification.

The above is intended to show that this is neither the right way to understand these cases, nor the right moral to draw from them. What is wrong with Norman and Truetemp

---

7 Even in the earliest printed statement of a reliabilist theory of justification (Goldman, 1979), we get a no-undermining clause. This is one of the reasons for thinking that nothing as simple as SR has made it to print.
is not that they lack an argument, for so do Nyrmoon and Normina, whose beliefs are justified (or at least are not obviously unjustified). Besides, Truetemp, in case (3), does have an argument, but this argument does not seem to help (even though it would suffice for the \textit{prima facie} justification of his perceptual beliefs). Since Nyrmoon’s and Normina’s beliefs are just as meta-incoherent as Norman’s, despite their being justified (or at least very nearly so), the so-called “meta-incoherence problem” fails to show that subjective rationality or internal justification, as glossed above, are necessary components of justification in general.

Perhaps these authors are correct in claiming that every justified belief requires some sort of inferential support -- that even all privileged beliefs are nonbasic, but the Norman and Truetemp cases do not provide an argument for this claim. To argue for this, we would need examples concerning perceptual beliefs or some other beliefs that we had antecedent reason to think were privileged beliefs. We have no such antecedent reason in the case of Norman’s clairvoyance or Truetemp’s temperature beliefs.

My view is that the “meta-incoherence problem” for simple reliabilism can simply be subsumed as a special case of a more general problem: that simple reliabilism implies that all beliefs are privileged. Thus, I propose to call this the “problem of privilege” instead of the “meta-incoherence problem”. This provides a more general and more accurate description of where simple reliabilism goes wrong. Viewed in this way, however, we see that the Norman and Truetemp cases do not show what they are intended to. Norman and Truetemp cases are \textit{supposed} to show that all beliefs require inferential support. What they \textit{in fact} show is that \textit{some} beliefs require inferential support from a generally adequate argument. This is a considerably weaker claim. So the problem with simple reliabilism is not necessarily that it allows belief in \textit{some} propositions to be justified in the absence of
inferential support — we have not yet seen an argument showing that there is anything wrong with this — the problem is that simple reliabilism allows belief in any proposition to be justified in the absence of inferential support. There are, to be sure, other problems with simple reliabilism, most notably the generality problem (Feldman, 1985) and the demon-world problem (Lehrer & Cohen, 1983; Cohen, 1984). Still, finding a solution to the problem of privilege would yield a markedly improved version of reliabilism, even if it is still subject to other difficulties.

1.2 The Problem of Privilege and Complications to Simple Reliabilism

As mentioned above, SR is less complicated than any version of reliabilism actually endorsed in print. Some of these more complex reliabilist theories might seem to offer some hope for dealing with the problem of privilege. Goldman’s earliest reliabilist theory (1979) distinguishes between belief-dependent and belief-independent processes. Alston (1988) requires that the grounds of a belief be accessible to the agent. Perhaps Goldman’s (1986) normal worlds analysis would improve on SR with respect to the problem of privilege. Even though all of these modifications yield improvements over SR, I want to argue in this section that none of these approaches solves, or even really addresses, the problem of privilege. The point, however, is not simply to chastise some philosophers for having gotten things wrong. The following is intended to illustrate the importance of the P/N distinction and the plausibility of the P/N thesis, while at the same time providing further argument for the above claim that the real import of Norman- and Truetemp-type examples is that they indicate that no adequate theory can deny the existence of nonprivileged beliefs. In addition, seeing which kinds of moves do not handle the problem of privilege will steer us in the direction of one that will. I cannot address every reliabilist
We can begin with the version of reliabilism expounded in Goldman’s (1979) paper, “What is Justified Belief?”:

(WJB): (1) a belief that is caused or causally sustained by an unconditionally reliable belief-independent process is *prima facie* justified,

(2) a belief that is caused or causally sustained by a conditionally reliable belief-dependent process is *prima facie* justified, and

(3) no other beliefs are *prima facie* justified.  

What makes this analysis of *prima facie* justification pertinent here is that it holds different kinds of beliefs to different standards. Beliefs that result from belief-independent processes (i.e., processes that do not take any beliefs as input) require unconditional reliability (i.e., a high objective probability of truth), whereas beliefs that result from belief-dependent processes (processes that do take beliefs as inputs) require only conditional reliability (high conditional probability of truth, given true belief inputs).

Although WJB distinguishes belief-dependent and belief-independent processes and the corresponding notions of conditional and unconditional reliability, this does not help to solve the problem of privilege, for neither of these two distinctions maps cleanly onto the distinction between privileged and nonprivileged beliefs. First of all, the distinction

---

8 In particular, I will not discuss any of Sosa’s writings in this section. The main reason for leaving him out is that, as far as I can tell, despite his focus on his reliabilist account of apt belief, he really intends for his generic coherentist account of justification to be taken seriously and to handle the meta-incoherence problem. This seems to be the view of his (1991a) and at least part of his (1991b). I have, however, already discussed the way in which I think that he misses the real import of the problem of privilege.

9 Goldman offers an account of undermining in order to get from *prima facie* justification to *ultima facie* justification. I omit the details here, since my concern is only with *prima facie* justification.
between belief-dependent and belief-independent processes fails to capture the distinction between privileged and nonprivileged beliefs. Although perception is both a belief-independent process and a source of privileged beliefs, memory is a paradigmatic instance of a belief-dependent process (Goldman, 1979), although memory beliefs, as we saw in Chapter 2, are also prime candidates for being privileged beliefs. In addition, the distinction between conditional reliability and unconditional reliability differs from the distinction between generally adequate inferential support and support by an otherwise substandard argument. It is possible for a process to be conditionally reliable in the sense of being reliable whenever it is given true beliefs as inputs, even though the conjunction of the inputs and the outputs of these processes constitutes nothing like an argument, let alone a generally adequate one.

Thus, the problem of privilege still arises. Whether we stipulate that Norman's clairvoyant belief is the result of a belief-independent process or a belief-dependent process, our intuition remains that his belief is unjustified. It is still a nonprivileged belief, a belief for which Norman needs, and lacks, a generally adequate argument. (In fairness to Goldman, the distinctions between belief-dependent and belief-independent processes and between conditional and unconditional reliability are almost certainly not intended to solve the problem of privilege, since WJB was formulated before any instances of the problem of privilege had been published.)

Another complication to simple reliabilism is offered by Alston in his "An Internalist Externalism":

(IE): A belief is justified if and only if it is based on an accessible ground, and this ground is a reliable indicator of the truth of the belief.

Of concern here is the requirement that the ground be accessible. Alston does not intend
for this to impose a general "subjective rationality" or "internal justification" constraint in
the sense mentioned above. The ground of the belief needs to be accessible, but the
adequacy of the ground does not. Does this relatively weak internalist constraint help to
solve the problem of privilege? No. Again we are confronted with a view that imposes too
few requirements on justification, thus having the result that too many beliefs are
privileged.

First note that IE cannot distinguish Norman from Nyrmoon, since the two are
supposed to be doxastically and phenomenologically identical. If the ground is accessible
for one of them, then it is accessible for the other, and we can suppose that not only are the
relevant processes equally reliable, but that the grounds on which Norman and Nyrmoon
base their beliefs are equally reliable indicators of the respective beliefs. So if we suppose
that both do use some accessible ground, a state of being appeared to in a certain way,
perhaps, then IE is too weak, for it wrongly implies that Norman is justified. If, on the
other hand, we suppose that neither has an accessible ground, then IE is too strong, for it
wrongly implies that Nyrmoon is not justified.10

The reason IE cannot distinguish Norman from Nyrmoon is that IE is not in a
position to distinguish two importantly different sorts of basing relations, which I will call
simply "privileged basing" and "nonprivileged basing". Nonprivileged basing, as I intend
the term, occurs whenever a belief is based on some ground in such a way that the belief is

10 There is obviously a more general issue lurking here. It seems possible for a
creature to be built in such a way that it has no percepts, no sensory experiences. Such a
creature might have perceptual systems otherwise like ours, except that its perceptual
systems do not deliver percepts or sensory states to consciousness but only belief states.
There is no reason to think that none of this creature's pure perceptual beliefs can be
justified, even though none of these beliefs are based on accessible grounds. It is in this
sense that IE seems too strong. Our perceptual beliefs tend to be based on sensory states,
but this does not seem to be a necessary feature of perceptual belief or of perceptual
justification.
inferred from that ground via some generally adequate argument. Privileged basing occurs whenever the belief is either based directly on that ground, or inferred via some other (otherwise) substandard argument. The problem of privilege concerns the basing relation, not the nature of the grounds. Alston's solution, since it focusses only on the nature of the grounds, cannot address this problem.

Looking at it this way, it becomes clear how we might generate further counterexamples to IE: find a belief that is intuitively nonprivileged and stipulate that it is based on an accessible ground that reliably indicates the truth of the belief in question, and stipulate further that the sort of basing in question is privileged basing. IE will wrongly imply that the belief is justified. Suppose that, for some strange reason, the apparent size of the moon in my visual field reliably causes me to form the belief that the moon is 2172 miles in diameter. We can suppose that this causal relationship realizes a basing relationship, although not one that involves the use of any generally adequate argument. This accessible ground (the apparent size of the moon in my visual field), because it reliably causes a true belief about the actual diameter of the moon, is a reliable indicator of the belief in question. Surely, however, my belief about the diameter of the moon is unjustified, and, it seems to me, no more justified than it would have been had it been caused by cosmic rays, directly, without the intervention of the accessible sensory state.\footnote{Plantinga (1993b, p. 191) offers similar counterexamples. However, as I will point out shortly, he misdiagnoses the problem.} This is a belief for which I need an argument, and having an accessible ground is not a satisfactory substitute for having an adequate argument. IE, like SR, fails at least partly because it implies that all beliefs are privileged.

Plantinga (1993a) seems to come closer than other reliabilists to noticing the
distinction between privileged and nonprivileged beliefs (although his topic is warrant, rather than justification). While discussing Alston's claim that only the grounds of a belief, and not their adequacy, need to be accessible, he insists that "there isn't anything at all like a simple, single answer to the question whether warrant for grounded beliefs requires that the subject know that the ground is an indicator of the belief; sometimes this is required and sometimes it is not" (p. 44). This claim can be modified to yield the following principle:

(P): The justification of some, but not all, beliefs requires that the believer be justified in believing that the ground is a reliable indicator of the belief in question.

This is almost the P/N thesis. Almost, but not quite, and as a result, the principle is too weak. The chief shortcoming of this principle is that it does not require a generally adequate argument, and, as such, runs afoul of the Audacious Induction of Chapter 2.

Suppose, as in Chapter 2, that I am justified in believing (and it is true) that my belief that the mass of the top quark is 175 GeV is a reliable indicator of the fact that the mass of the top quark is 175 GeV. My belief is still unjustified, because the Audacious Induction is not a generally adequate argument.

One final complication to simple reliabilism to be considered in this section is Goldman's (1986) normal worlds reliabilism, according to which it is not the actual reliability of a process that determines whether the belief is justified, but the reliability in "normal worlds"; i.e., worlds that are, in general, the way we think the actual world is. Goldman has since disavowed this position (see, e.g., Goldman, 1988; 1992a), and even when he held it, he did not take advantage of it in replying to Norman-type examples. (In section II, I will discuss his (1986) response to these sorts of examples.) Nonetheless, the

---

12 Goldman's actual proposal is that a rule system is right if and only if it licenses a sufficiently high ratio of true beliefs in normal worlds. The difference between these two proposals, however, will not affect the present point.
version of reliabilism that I find most promising bears important affinities to this normal worlds approach, and it will thus be helpful to see the extent to which a normal worlds reliabilism can handle the problem of privilege.

Although I am not aware of anyone who explicitly takes this approach, it could be argued that clairvoyance is not reliable in normal worlds. In normal worlds (although perhaps not in the actual world or in Norman’s world) there simply are no genuine parapsychological phenomena. The physical laws that obtain in normal worlds seem to preclude the sorts of causal relations clairvoyance requires. This hypothetical response, I think, fails, but it does so for an interesting reason. Once more we have a view that cannot distinguish Norman from Nyrmoon. If clairvoyance is unreliable in normal worlds, then Nyrmoon ought to be in the same epistemic position that Norman is in; Nyrmoon’s clairvoyance beliefs (or anybody else’s) ought to be unjustified. Either clairvoyance is reliable in normal worlds, or it is not; it cannot be reliable-for-Nyrmoon-in-normal-worlds but not reliable-for-Norman-in-normal-worlds. The beliefs that serve to delineate the set of normal worlds are supposed to be general beliefs and not about particular individuals (Goldman, 1986, p. 107). Thus, the normal worlds analysis cannot get the Norman case right without getting the Nyrmoon case wrong.\textsuperscript{13} The normal worlds approach does not help with the problem of privilege, for it does not require of any beliefs that they are supported by a generally adequate argument. Even though it might offer a response to the Norman case, it does not address the problem of privilege in its full generality.

Still, there seems to be something right about a normal worlds component to a theory of justification. Recall that one of the morals of the comparison between Norman

\textsuperscript{13} Goldman (1988) uses a similar example in the course of explaining his misgivings about the normal worlds approach.
and Normina and Nyrmoon is that it must be our beliefs about the relevant processes, rather than the agents' own beliefs, that accounts for the difference in our judgments concerning the justifiedness of their beliefs. Norman and Nyrmoon are doxastically identical, by hypothesis, yet one is justified and the other is not. Yet it is not our beliefs about the reliability of clairvoyance that account for the difference in our intuitions in these two cases, for we only have one belief about the general reliability of clairvoyance. Instead, I will argue that it is some other beliefs of ours, beliefs about some psychological properties that distinguish Norman from Nyrmoon, that distinguish Norman from Nyrmoon, and the way that these beliefs underwrite the distinction is by ascribing privileged status to Nyrmoon’s belief but not to Norman’s. This argument will begin in Section 3 below, but it will occur primarily in Chapter 4.

2. Reliabilism and Nonprivileged Beliefs

The foregoing is intended to illustrate the robustness of the problem of privilege and to suggest that any adequate theory needs to countenance nonprivileged beliefs. Although the proposals examined above may stave off certain instances of the problem of privilege, none address the problem in its full generality. Some beliefs must be held on the basis of a generally adequate argument in order to be justified. Barring any serendipitous accidents, no complication to simple reliabilism is going to adequately handle the problem of privilege without explicitly recognizing this. I have not run through all the various twists and turns reliabilists have proposed over the last twenty years. But this would not be necessary; it is clear enough that the problem of privilege will not go away until we can formulate a version of reliabilism according to which, being held on the basis of a generally adequate argument is necessary for the prima facie justification of nonprivileged beliefs.
2.1 Ex Ante Undermining

As mentioned above, Goldman's treatment of the Norman case in his (1986) book does not rely on the normal worlds approach. Here Goldman appeals instead to a no-undermining condition, but it is an unusual kind of undermining that is at issue. Norman's belief is not undermined by his current cognitive state, by hypothesis, but it is undermined, Goldman suggests, by a belief that he would have if he were following his doxastic obligations (the belief that his clairvoyant powers are not reliable). "Norman ought to reason along the following lines: 'If I had a clairvoyant power, I would surely find some evidence of this. . . . Since I lack any such signs, I apparently do not possess reliable clairvoyant processes'" (p. 112, italics in original). This response, if it works, would solve the problem of privilege and do so in a way that is consistent with denying the P/N thesis.

How? Take any belief that is not held on the basis of a generally adequate argument, and it is possible to claim that the agent is epistemically obligated to believe that the process that produced that belief is an unreliable one, in accordance with the argument Goldman offers above. Goldman could thereby use this approach to argue that the beliefs I have been calling nonprivileged really are privileged in the sense of being prima facie justifiable in the absence of any argument; it is just that, due to some epistemic obligations, the beliefs are always undermined and thus never ultima facie justified. This would capture the relevant intuitions without admitting the P/N distinction.

Positing such epistemic obligations, however, brings us to the edge of a slippery slope. Are children obligated to reason as follows, simply because they are capable of doing so reliably?
If my perceptual processes are reliable, then surely I would have some evidence for this. Thus, if I cannot find some indication of the reliability of my perceptual processes, I shall be forced to conclude that they are unreliable. This seems unduly stringent, especially for an externalist. How are epistemic obligations supposed to fit into the general story? Is epistemic permission a matter of truth ratio of \( n \), while epistemic obligation is a matter of truth ratio of \( n + m \)? It seems unlikely, since the psychological processes that subserve my elementary mathematical abilities seem to be very highly reliable. Am I epistemically required to spend my free time engaged in addition and subtraction? And if it is not reliability that makes for obligation, we deserve to be told what it is.

Goldman is better off avoiding the appeal to obligation altogether. The rejoinder to the Norman case that he seems to have in mind here might just as easily be captured by claiming that, obligation or no, Norman is \textit{ex ante} justified in believing that he has no reliable clairvoyant powers and this suffices to undermine the belief that the president is in New York undermined. But I can see no reason to think that this is necessarily the case. Is it truly impossible for Norman to have no evidence for or against the reliability of his clairvoyance? Goldman seems to think so: he claims of BonJour’s Norman case that “it is hard to envision this description holding” (p. 112). But \textit{ex ante} justification is presumably, like \textit{ex post} justification, a matter of reliability (this view was explicitly endorsed in his [1979]), and reliability is a decidedly contingent feature of cognitive processes. There must be some world in which every process that would produce the argument that Goldman requires of Norman is unreliable, and even in these worlds, Norman strikes us as

---

14 I presume that one is epistemically obligated to believe that \( p \) only if one is (at least \textit{ex ante}) \textit{justified} in believing that \( p \).
Another significant problem with the *ex ante* justification response, in my view, is that it seems to imply that Nyrmoon cannot be justified either. Since Nyrmoon is doxastically identical to Norman, if being *ex ante* justified in believing the conditional ‘if I have reliable clairvoyant powers, then I have some evidence for it’ suffices to undermine Norman’s belief, it ought to undermine Nyrmoon’s belief as well (assuming that Nyrmoon’s *modus tollens* processes are also sufficiently reliable). But, intuitively, Nyrmoon’s belief is justified. Worse yet, the same seems to hold for Normina, the unreflective human perceiver, for presumably she would be justified in believing an analogous conditional. Perhaps this general approach can be salvaged, although it does not look very hopeful.

There is one way that I can see to get *ex ante* justification to provide a response to this class of counterexamples. We could develop an account of epistemic obligations which does not rely on the mere reliability of the process in question (for there are too many difficulties with that approach) and does not apply so generally that it is impossible for an agent to have no evidence one way or the other concerning the reliability of their faculties. By invoking the P/N thesis, we might find some grounds on which to claim that Norman is flouting an epistemic duty while Nyrmoon and the children who fail to reason as above concerning their perceptual beliefs are not. But if we’ve already accepted the P/N thesis, such a response is unduly complicated. Accepting the P/N thesis allows for a much simpler solution to the problem of privilege: the reason Norman is unjustified is that he does not have an argument, and the belief in question is a nonprivileged belief. This seems more parsimonious than positing epistemic obligations and no further from the basic spirit of externalism. All told, the *ex ante* justification gambit looks unpromising. Better to
explicitly endorse the P/N distinction and see what can be done from there.

2.2 Restricted Externalism

The fact (or at least so I take it) that any adequate epistemology has to count some beliefs as privileged and some beliefs as nonprivileged suggests a surprisingly simple way to develop a version of reliabilism that avoids the problem of privilege. In particular, we might reformulate SR or something like it as a thesis only about the justification of privileged beliefs. And, of course, there is no reason why one could not also claim that reliability is what distinguishes good inferential practices from bad ones. We are guaranteed to have a reliabilist theory that avoids the problem of privilege if we can develop a version of reliabilism that is sensitive to the privileged/nonprivileged belief distinction, a theory we might call "P/N reliabilism". Something like the following could serve as a first approximation to a P/N reliabilist theory.

(P/NR): (1) A privileged belief is *prima facie* justified if and only if it is produced by a reliable process.

(2) A nonprivileged belief is *prima facie* justified if and only if it is produced by a conditionally reliable inferential process that takes as its ultimate premises justified privileged beliefs.

I am just trying to get the basic ideas out on the table, and as such, I am ignoring certain complexities. There will almost certainly need to be further constraints, e.g., on the sort of inferential process referred to in (2), on the kind of reliability at issue in (1), and so forth. Still, the basic idea seems appealing enough, and it appears to handle the problem of privilege without raising further difficulties for reliabilism. There is, of course, one new difficulty, which is that of providing a theory of which beliefs are privileged and which
beliefs are nonprivileged. But since, as I have been arguing, all the major internalist theories are already committed to the privileged/nonprivileged distinction, this is everyone's problem and thus not really a difficulty for reliabilism.

This view bears an important resemblance to something that BonJour (1985) has criticized under the heading of "restricted externalism": "an obvious and initially appealing response for the externalist would be simply to pull in his horns, abandon the unnecessarily general form of his view discussed above, and advocate it only as restricted to the range of cases which are his main interest" (p. 49). One of the main problems for such a view, according to BonJour, is that the restriction is unprincipled: "Having in effect accepted [the Norman] argument as applied to nonstandard cases (for otherwise why restrict his position?), the restricted externalist must explain clearly why it does not apply equally well to the more familiar cases with which he is concerned" (p. 50).

The P/N distinction gives us an easy answer: we have antecedent reason to think that privileged and nonprivileged beliefs are going to have to be treated differently. This restriction is not unprincipled, or if it is then this is because there is something unprincipled about our epistemological intuitions, and any theory that takes these intuitions as its explicanda is going to have to learn to live with it. The P/N thesis (and a fortiori the P/N distinction) figures too heavily into our pretheoretical epistemological intuitions to be ignored. BonJour himself tacitly relies on the P/N distinction, as I argued in Chapter 2. He must do so if he is going to retain even the outlines of his account of perceptual justification while at the same time denying the legitimacy of the Audacious Induction. P/N reliabilism is no worse off than any other theory for relying on the P/N distinction. Internalism, as I argued in Chapter 2, does not provide an alternative to this "unprincipled" distinction, for the internalists agree: some beliefs require generally adequate arguments and
others do not.

I do agree with BonJour on two points. First, we need some account of this distinction; that is, we need a set of necessary and sufficient conditions for a belief's being privileged. In fact, this is going to be the subject of the next chapter and is one of the central goals of the larger project undertaken here. Yes, some account is needed, although it is not just restricted externalists who need it. Secondly, I agree with BonJour that there is something slightly uncomfortable about simply tacking the P/N thesis onto an existing account of epistemic justification. It would be nice if we could develop a more general theory of justification into which the P/N thesis would fit naturally.

3. Two-Component Reliabilism

In this section, I try to elucidate Goldman's most recent (1992a, 1994) theory of epistemic justification and show how the P/N thesis might be incorporated into such a view. I concentrate on this particular theory partly because I find it to be the most independently plausible reliabilist theory currently on offer and partly because it provides an opportunity to discuss the descriptive epistemological project (which figures heavily into Chapter 4) in more detail. Goldman has already argued for this view, and I will not attempt to add much here, though I will point out some of the virtues of this theory in what follows. My intent here is not to defend this theory but to improve it by making it explicitly accommodate the P/N thesis.

Before rehearsing Goldman's account, it is important to review the mission of descriptive epistemology, discussed in the first chapter. Descriptive epistemology, again, is the task of providing a set of concepts and principles that best matches our folk understanding of justification and/or knowledge, to account for our intuitive judgments of
justifiedness and/or knowledge. This is the task I have been concerned with so far, and it will occupy me in this and the following chapter as well.

Goldman's (1992a) Two-Component Reliabilism (henceforth, TCR) is explicitly aimed at the descriptive project; it is an attempt to provide a framework for systematizing our intuitions concerning epistemic justification.\(^{15}\) If we really take seriously the distinction between descriptive epistemology and normative epistemology, it seems clear that the task of descriptive epistemology is simply the task of characterizing the "knowledge structures" that underlie our first- and third-person ascriptions of knowledge and/or justification. The phrase 'knowledge structures' is used here in the psychologist's sense, and as such does not itself imply anything about truth or justification, or even, strictly speaking, belief. A knowledge structure is simply a set of representational, or informational, states, which is operative in the performance of some task, typically a task that requires "intelligence" of some sort. These states are belief-like in that they encode information about the world, but they might not have all the properties that are required in order for something to count as a belief.\(^{16}\)

Goldman's account of our intuitions about justifiedness begins with a fairly standard psychological theory of concepts. According to both prototype and exemplar theories, subjects classify targets as instances of some category according to the degree of

\(^{15}\) I focus on the (1992a) version, although it is also defended in his (1994), simply because the former is more thoroughly explicated. As far as I can see, there are no substantive changes between the two papers. Also, I use his suggested name, "Two-Component Reliabilism" (1992a, p. 315) rather than "Virtue Reliabilism", which he also suggests (p. 315), in order to avoid any confusion with the rather different view espoused by Sosa (1991a; 1991b).

\(^{16}\) For example, Ramsey, Stich, and Garon (1990) cite functional discreteness as a necessary condition for something's being a belief, and it is likely that knowledge structures lack this property.
similarity that obtains between the target in question and some representation of the category in question.\textsuperscript{17} Goldman suggests that we categorize certain beliefs as justified or unjustified based on our understanding of the process that produced them. We have stored representations of both virtuous and vicious belief-forming processes, and a target belief is categorized as justified if (and to the extent that) the process thought to produce it resembles one of the processes on the attributor's virtue list. If the process thought to produce the belief in question closely resembles one of the processes on the attributor's vice list, on the other hand, the belief is classified as unjustified.

Processes are placed on the virtue or vice lists, according to this view, on the basis of (perceived) reliability; processes that are deemed reliable, like perception, deduction, memory, and so forth, end up on the virtue list and unreliable processes like wishful thinking, ignoring relevant evidence, and the like, end up on the vice list. These knowledge structures (the virtue and vice lists), like most knowledge structures, are resistant to capricious change. Thus, being told that, say, Norman's clairvoyant powers are completely reliable does not cause us to add clairvoyance or even Norman's clairvoyance to the virtue list (although perhaps being actually convinced for a long time that clairvoyance is generally reliable would).

I want to be explicit about how I understand this view. Here are what I take to be the essential features of TCR:

(1) It takes quite seriously the strictures of descriptive epistemology and insists that

---

\textsuperscript{17} The difference between prototype and exemplar theories lies mainly in the kinds of representations they posit. Prototype theorists, in the spirit of Locke, see concepts as abstract, unitary representations that explicitly encode central tendencies of the members of the category. Exemplar theories, on the other hand, more closely resemble the views of Berkeley and Hume. On a generic exemplar theory, category representations are sets of representations of the individual members (exemplars) of the category in question. For a nice discussion, see Smith & Medin (1981).
the proper way to pursue the descriptive project is to elucidate, in a general way, the content of the knowledge structures that underlie our intuitive judgments about justifiedness.

(2) It sees these judgments as being made on the basis of the psychological process that produced or sustains the belief in question.

(3) It claims that the feature that distinguishes vicious from virtuous processes is (perceived) reliability.

Regarding (1), I see Goldman's claims about the structure of concepts (the prototype theory) to be largely a matter of convenience; if this theory of concepts turns out to be false, it will not seriously damage Goldman's position.

TCR improves on earlier reliabilist theories in several respects. One of the virtues of this view, as I see it, is that it incorporates the advantages of a normal worlds account without inheriting its disadvantages. It is judged reliability that determines which list a process ends up on. Thus, S's firm and longstanding (though perhaps unconscious) conviction that perception is reliable will cause S to place perception (in certain conditions) on S's virtue list, whether perception is in fact reliable or not. It is in this sense that TCR is similar to the normal worlds analysis of Goldman (1986). There are important differences, however. First, it is not our beliefs that determine whether a particular belief is justified, but a specific attributor's beliefs that determine whether that attributor judges a particular belief to be justified. Second, it is not even all of an attributor's beliefs that determine this judgment, but only those that are embedded in the knowledge structures that subserve the capacity for forming such intuitive judgments of justifiedness.18

TCR also solves two of the most serious problems facing reliabilist theories in

---

18 This second point will receive more detailed treatment in Chapter 4.
general, the demon-world problem and the generality problem. It solves the demon-world problem by adopting what is essentially a normal worlds account: perception and the like are thought (by the typical attributor) to be reliable, and therefore they are on the (typical attributor's) virtue list. The demon's victims are judged to be justified because the processes they are using match processes we take to be reliable, even though these processes are not reliable in the demon world. This approach, like that of Goldman (1986), but unlike that of Sosa (1991a, 1991b), works even if it turns out that the actual world is a demon world.

TCR's solution to the generality problem is less obvious. Very briefly, the generality problem (Feldman, 1985) is the problem of specifying which process type (or more generally, which process type-environment type pair) is going to count in determining process reliability. If the process that produced my belief that \( p \) at time \( t \) is specified too broadly, e.g., as an instance of cognition, then we lose the ability to distinguish justified (or more justified) from unjustified (or less justified) beliefs, since all beliefs are the result of cognition. On the other hand, if the relevant process type is specified too narrowly, there will only be one token of the process type: the process token that caused my belief that \( p \) at time \( t \). But then it would seem that the reliability of the relevant process would have to be either 1 or 0, depending entirely on whether the belief that \( p \) at time \( t \) is true or not. Any process type that is neither too broad nor too narrow would seem to be chosen merely to get the right answer; there does not seem to be any principled way to specify the relevant process type at an intermediate level of generality.

---

19 In fact, Goldman seems to be discussing this in his (1992a, pp. 161-2) but even there he is not explicit about this being the generality problem, so perhaps he has something slightly different in mind. The discussion here will elaborate on, and diverge from, his in certain respects.
I claim that TCR has a solution to this problem, although perhaps 'solution' is a bit too strong. More accurately, on a two-component reliabilist theory, the generality problem becomes a problem of detail rather than a problem of principle. This is because TCR does not have to specify the relevant process type. Because it takes the strictures of the descriptive project so seriously, the explicandum of TCR is not so much justification as attributions of justification. The question about what level of generality of process type determines the reliability of the process becomes the question of how narrowly individuated the process is that the attributor takes to be relevant, and this is just the question of how fine-grained the attributor's virtue and vice lists are and under what description the attributor conceives the target belief. The relevant process type is the one that best matches the target belief, and this is going to depend on how the attributor conceives of the target belief as well as how many and which processes (or process-environment pairs) are represented on the virtue and vice lists.

Certainly, attributors distinguish between different kinds of perception and perception in different environmental conditions. This suggests that the virtue and vice lists contain representations that are fine-grained enough to distinguish, e.g., vision in several kinds of environmental circumstances. But just how many distinctions the virtue and vice lists contain and just where these lines are drawn is clearly a matter of detail rather than principle, from the perspective of TCR. TCR is no more obligated to answer this than to provide a list of attributors and the processes that each thinks vicious or virtuous. If there is anything unprincipled about the way that attributors individuate the relevant process types, then this is at most something for which we should blame the attributors, not TCR.

3.1 The PIN Distinction in Two-Component Reliabilism
How does TCR, thus understood, handle the P/N thesis and the problem of privilege? So far, it does not. My complaint here should be familiar by now: TCR cannot yet predict or explain the difference between Norman and Nyrmoom. The view just adumbrated does not explain why I find Nyrmoom to be justified and Norman to be unjustified, for it seems clear that either I do or do not have clairvoyance on my virtue list, especially given that the virtue list is not subject to quick and temporary change. Thus, I want to claim that the reason why we take Norman to be unjustified is not (or at least not merely) that we do not have clairvoyance represented as a virtue. This is not to say that TCR gets the wrong answer here, for recall that, according to TCR, it is similarity to some prototype on the virtue list that yields judgments of justifiedness. Clairvoyance itself need not be on the virtue list, so long as it is sufficiently similar to something that is. TCR could handle both of these cases at the same time if it could find some principled grounds on which to claim that Nyrmoom’s clairvoyance is more similar to some virtuous process than Norman’s clairvoyance is.

This raises a more general question about the nature of similarity. Presumably, we are capable of forming judgments about beliefs that result from novel processes which have not yet been thought of and placed on a virtue or vice list, and it is because of this that some targets must be categorized on the basis of similarity rather than outright matching. But if similarity is to play such a crucial role in the explanation of epistemic attribution, we need some account of what similarity is. TCR is not obligated to tell us what processes are represented on whose virtue lists, but it does owe us a sketch of what similarity amounts to. The process that produces Truline’s belief shares infinitely many properties with the process that produces my visual belief that there is something white in front of me. Both processes are processes; neither process is a lawnmower; neither is identical to the number
7 or the number 8 or the number 9, etc. Clearly we need some specification of which sorts
of properties are relevant for the purposes of categorization.

Similarity is a notoriously slippery notion, and it is especially troublesome in the
present context of determining category membership. This is a problem in the concept
literature in general, not just concerning the use to which Goldman wants to put a theory of
concepts. As many have pointed out (Quine, 1960; Armstrong et al., 1983; Murphy &
Medin, 1985), it is hard to find a non-circular specification of similarity available for use in
concepts. Making good on TCR requires some independent specification of similarity; we
need some assurance that target processes are deemed virtuous because of their similarity to
stored prototypes, and not that they are deemed similar because of their shared virtue. This
is true of any theory that attempts to account for our ability to categorize things.

Now the standard solution in the psychology literature to this problem of similarity,
and the only one I am aware of, is to endorse a theory-theory of concepts (Murphy &
Medin, 1985). According to the theory-theory of concepts, subjects use, perhaps in
conjunction with prototypes or exemplars, a host of general world information. For
example (see Murphy & Medin, 1985), if told that an object is three inches in diameter and
asked whether it is more likely to be a pizza or a quarter, we are pretty confident that it is
more likely to be a pizza. Although the only attribute we are given here is size, and three
inches is closer to the size of a typical quarter than to the size of a typical pizza, our
background knowledge about the world presumably includes the belief that pizzas are likely
to vary in size and quarters are not.

So according to the theory-theory of concepts, our general theory about the world
constrains the use of features in similarity judgments. The background theory is thought to
determine which similarities are relevant in a given context. Certainly this view needs to be
worked out in some more detail for the purposes of psychology. For the present purposes, however, it is enough to note that we can flesh out a brute appeal to similarity by specifying the pertinent parts of the subject’s general background theory. In this way, we can make vague talk about similarity relatively precise and predictive.

The problem of privilege, and in particular, the difference between Norman and Nyrmoon, suggests that something like the following is encoded in the knowledge base we use in forming intuitive judgments about justifiedness:

Nyrmoon’s belief results from a different kind of process than Norman’s belief. The process that Nyrmoon uses has some property, $P$, which the process that Norman uses does not, and $P$ is a property such that all and only processes that have $P$ produce privileged beliefs.

In the spirit of TCR, I suggest that the P/N distinction maps onto a distinction among kinds of processes (this will receive further defense and clarification in Chapter 4). What the property $P$ is that distinguishes privileged from nonprivileged beliefs or processes is the subject of Chapter 4.

The intuitive truth of the P/N thesis indicates that the typical attributor will have two kinds of processes on her virtue list: processes that are thought to be reliable and to produce privileged beliefs and what we might call “argumentational processes” (i.e., processes the execution of which instantiates generally adequate arguments that are in the relevant sense possessed by the subject) that are thought to be reliable. These two should be exhaustive, since any non-argumentational process that delivers nonprivileged beliefs will deliver unjustified beliefs and would thus be considered vicious. Now the vice list might include some argumentational processes that are thought to be unreliable, and it might include even some processes that are thought to produce privileged beliefs, but which are also thought to
be unreliable.\textsuperscript{20}  

Now, TCR can claim that the reason Nyrmoon's belief is judged to be justified and Norman's is not is that the property $P$ is given a lot of weight in determining whether a target process is sufficiently similar to one already on the virtue list. One of the most important features of perceptual and memory processes, at least in the context of epistemology, is that they yield privileged beliefs. Whatever features it is in virtue of which they do this ought to be given a very high weight in determining whether some other process is similar to them or not. It is (at least largely) because Norman's belief is the result of a process that does not have $P$ that it is not judged to be sufficiently similar to such privileged processes as perception and memory. And it is (at least largely) because Nyrmoon's belief is the result of a process that does have $P$ that this process is judged to be sufficiently similar to perception. This view requires that TCR add some specification of which shared properties are most relevant for the purposes of matching, but I have argued that this would be necessary anyhow, at least for a fully developed theory.

Obviously, this account will need more detail in order to become a full-blown epistemological theory, but I will leave it unfinished for now, since my main concern at this point is merely to show how a reliabilist might find a way to fit the P/N thesis comfortably into the larger epistemology. Again, we are left with the pressing question of what this property is that distinguishes privileged and nonprivileged beliefs. Before turning to this topic, however, I want to examine the relation between the present version of TCR and the sketch of P/NR from Section 2, above.

\textsuperscript{20} Recall that the characterization of privileged beliefs from Chapter 2 is neutral with respect to what is required for the justification of a privileged belief, except that having a generally adequate argument is \textit{not} required. Privileged beliefs might very well still require a particular causal history (e.g, being the result of a reliable cognitive process) for their justification.
3.2 Two-Component Reliabilism and P/N Reliabilism

Recall the outline of P/N reliabilism given above:

(P/NR)

(1) A privileged belief is *prima facie* justified if and only if it is produced by a reliable process.

(2) A nonprivileged belief is *prima facie* justified if and only if it is produced by a conditionally reliable inferential process that takes as its ultimate premises justified privileged beliefs.

Although there are obvious affinities, this is not quite the same as the version of TCR developed above. First of all, TCR was concerned with judgments or attributions of justifiedness, rather than with justification itself.

More importantly, according to TCR, even perceived reliability is neither necessary nor sufficient for judged justification. This is partly because, by adopting a non-definitional theory of concepts, TCR leaves very little if anything necessary or sufficient for anything. However, even though a non-definitional account of concepts is almost universally accepted among psychologists (and for good reasons, I think), it is not an essential feature either of TCR in general or of the version of TCR I have tried to develop. The more essential reason why perceived reliability is neither necessary nor sufficient for judged justifiedness, according to TCR, is that processes that are perceived to be reliable might still more closely match vices than virtues, and processes that are perceived to be unreliable might still more closely match virtues than vices.\(^{21}\)

---

\(^{21}\) For these reasons, one might wonder whether TCR is really a process reliabilist theory; cf. Markie (1996).
P/NR was developed specifically in order to handle the problem of privilege. TCR, or at least the version I advocate above, solves not only this problem, but the generality problem and the demon-world problem as well. Furthermore, by focusing on attributions of justification rather than justification itself, TCR promises a far superior approach to answering the questions of descriptive epistemology. There is something appealing about P/NR, as formulated above, but perhaps this sort of view should be saved for the normative project. One of the main reasons why it is impossible to translate TCR into P/NR is that the latter insists that reliability is necessary, though not sufficient, for justification. Denying the sufficiency of reliability helps to avoid the problem of privilege, but maintaining the necessity of reliability runs afoul of the demon world problem. Although I share the intuition that the victims of the demon are justified in their beliefs, I cannot help but think that there is a sense in which, if we were (or are) victims of such a demon, we ought to refrain from forming any beliefs. This might very well be the sense of 'ought' relevant to normative epistemology.

Despite its virtues, there is one stark limitation on TCR: TCR is specifically designed to provide a descriptive epistemological account of justification and not a normative account. The attributor plays an essential role in the formulation of TCR, but a normative epistemological theory is supposed to take justification as its target, not ascriptions of justifiedness. TCR does not seem to offer any very promising foothold here, for it is far from clear what would even be left of TCR once all reference to the attributor were eliminated. It is here that P/NR might be brought in; although it needs work, it seems fairly well-situated to play the role required in a normative epistemology.

What P/NR and TCR have in common is that both deny that reliability is sufficient for justification of all beliefs (or even that perceived reliability is sufficient for judged
Reliability is invoked, primarily, as sufficient for the justification of perceptual and memory beliefs and other beliefs that share the privileged character that makes reliability sufficient for the justification of these. The version of TCR developed above was able to capture this merely by adding some detail to the existing TCR framework, detail concerning the typical contents of the attributors' lists and the similarity weighting employed by the typical attributor. Every epistemology needs to incorporate the P/N thesis or risk either skepticism (by denying the existence of privileged beliefs) or admitting too many beliefs as justified (by denying the existence of nonprivileged beliefs). The version of TCR sketched above seems to avoid both unpalatable extremes and accommodate the P/N thesis in a natural and principled way.
CHAPTER 4:
AN ETIOLOGICAL, FACULTY-ORIENTED ACCOUNT OF THE P/N DISTINCTION

Recall that the privileged/nonprivileged distinction is one concerning the kind or amount of inferential support a belief requires for its justification. Privileged beliefs can be \textit{prima facie} justified on the basis of an otherwise substandard argument (and no argument at all is a limiting case of this), whereas nonprivileged beliefs require the support of a generally adequate argument. I have argued that the major internalist theories tacitly rely on this distinction and that the major externalist theories need to. This means that any epistemological theory owes us an account of which beliefs are nonprivileged and which beliefs are privileged, i.e., which beliefs require a generally adequate argument and which do not. I have characterized the P/N distinction in epistemic terms, but we also need an account that tells us, for any given belief, which set of epistemic standards it is subject to. I have argued that perceptual and memory beliefs are privileged, but it would be nice to know which (if any) other beliefs are also \textit{prima facie} justifiable on the basis of an otherwise substandard argument. In order to say anything more about which beliefs are which, we will need an account of the distinction in non-epistemic terms, or at the very least, one that does not hinge on the type of inferential support a belief requires for its \textit{prima facie} justification.

The purpose of this chapter is supply the tentative first steps toward such an account. Having only just now, in the previous two chapters, made the P/N distinction explicit, I do not suppose that the ensuing account will be anything like the final word on the subject. My primary goal here is to defend a certain framework for what I think would be an adequate account of the P/N distinction, a list of general features a theory accounting
for the P/N distinction should have.

In particular, I defend an etiological, faculty-oriented account of the P/N distinction, in Section 1. Since it is really a family of possible accounts that I am mainly concerned to defend, I will not say anything very precise there. Although my primary concern here is (and has been) with the descriptive epistemological project (mentioned already in Chapters 1 and 3), this will become much more explicit in Section 2, where I offer a particular kind of etiological, faculty-oriented, descriptive epistemological account. Again, however, it is the more general family of accounts that is my primary concern, and I although I think that the more specific theory described and defended in Section 2 has some plausibility, I offer it largely to illustrate the potential of the more general theory. In Section 3, I explore the beginnings of a normative epistemological account which takes as its starting point the descriptive account developed in Section 2. The accounts sketched in Sections 2 and 3 are rather tentative, and in Section 4 I discuss some of the reasons for this.

1. The General (Family of) Account(s)

Just as we have pretheoretical intuitions about which beliefs are justified and which beliefs are unjustified, we also have pretheoretical intuitions about which beliefs are privileged and which beliefs are nonprivileged. It is just that the latter intuitions are a little harder to recognize. The use of these intuitions in the previous two chapters, however, indicates that we do have them. It would be easier to access these intuitions about privilege were everyone to have the sort of intuitions that motivate physical object foundationalism.¹ For then, we could simply stipulate that an agent has no inferential support at all for the belief in question and no undermining evidence and read privilege off of the resulting

¹ Physical object foundationalism (POF) was defined in Chapter 2, p. 34.
intuition about the justification of that belief. If our intuition were that the belief was justified, then we would count it as privileged, and we would count it as nonprivileged otherwise.

This was the general approach of Chapter 3. Even those who would not think that the belief was justified would be able to see that the physical object foundationalist (and the reliabilist in particular) would not be embarrassed by the claim that the belief was justified. What mattered was not that the reader had the intuition that the belief was justified but only that the belief had the same epistemic status as perceptual beliefs of an unreflective agent. By now I think that the reader should have reasonably clear intuitions about whether or not a particular belief is privileged. Occasionally, however, I will return to the above approach in order to ensure that the intuitions are clear.

1.1 Cases of Privileged and Nonprivileged Beliefs

As argued in Chapter 2, some of the most familiar and paradigmatic cases of privileged beliefs are perceptual beliefs and memory beliefs. When I look at the round table in front of me and form the belief that there is something round in front of me, this strikes us, intuitively, as a belief whose prima facie justification does not require the support of a generally adequate argument, in the sense described in Chapter 2. Memory beliefs seem to be the same: when I believe, on the basis of memory, that the cat was under the desk, this is the sort of belief that can be prima facie justified on the basis of an otherwise substandard argument.

Some paradigmatic instances of nonprivileged beliefs also arose in Chapter 2. Our scientific or philosophical beliefs, for example, are generally not privileged. If I am sitting in my laboratory, and it suddenly strikes me that the mass of the top quark is 175 GeV, this
belief is not something that could be even *prima facie* justified by anything short of a generally adequate argument; no appeal to my own reliability or trustworthiness would suffice, nor would an attempt to explain the occurrence of the belief by appeal to the truth of the belief. Similar verdicts held for my belief that nominalism is true, or that it will snow a lot this winter.

A descriptive epistemological account ought to capture our intuitions concerning not just actual cases of privileged and nonprivileged beliefs, but concerning every possible case as well, and in order to ensure that the account will do this, we have to find a number of less ordinary cases of privileged and nonprivileged beliefs. In Chapter 3, we saw a number of these, where I argued that the Norman and Truetemp counterexamples to simple reliabilism were good counterexamples *because* Norman’s and Truetemp’s beliefs were not privileged. Norman is a perfectly reliable clairvoyant (although he has no (justified) beliefs for or against this claim), and Truetemp has had a device implanted in his head which produces reliably accurate beliefs about the ambient temperature (although he has no idea that the device is there). Intuitively, neither Norman’s clairvoyance belief nor Truetemp’s temperature belief is privileged.

However, each of these agents has an alien counterpart for whom the analogous belief *is* privileged. Nyrmoon is an alien with an innate clairvoyance module, which functions psychologically just like perception does for us. He is so unreflective, however, that, like Norman, he has no (justified) beliefs for or against the claim that he is a reliable clairvoyant. Vipertemp has a temperature sensing organ, just like Truetemp’s, except that Vipertemp’s organ is the result of millions of years of natural selection, rather than surgery. In each case we can suppose that the alien is doxastically and phenomenologically identical to his human counterpart, yet we have different epistemological intuitions about the aliens.
than we did about the humans. Unlike Norman and Truetemp, Nyrmoo and Vipertemp seem to be in the same epistemic position as Normina, an unreflective human perceiver (whatever one may think of her epistemic position -- whether her perceptual beliefs are justified or not). Since Nyrmoo’s clairvoyance beliefs and Vipertemp’s temperature beliefs have the same status as Normina’s perceptual beliefs, and since perceptual beliefs are privileged, if follows that Nyrmoo’s and Vipertemp’s beliefs are also privileged.

Once more, a belief’s being privileged does not imply that it is justified, or even *prima facie* justified. If the belief requires some sort of inferential support from other beliefs (à la MSF and coherentism) and is lacking it, then it will not be (even *prima facie*) justified. It may very well be that Norman, Normina, and Nyrmoo are all unjustified in holding the relevant beliefs, even though these beliefs still have different standings vis-à-vis privilege. And since it is privilege, rather than justification *per se*, that concerns me, it does not matter whether the reader joins me in thinking that Normina’s and Nyrmoo’s beliefs are justified. The point is only that Normina’s and Nyrmoo’s beliefs are privileged: that these beliefs *could be* justified, without the agent’s having a generally adequate argument for them.

The fact that Norman and Nyrmoo are doxastically and phenomenologically identical, even though Nyrmoo’s belief is privileged and Norman’s belief is nonprivileged places important constraints on any account of the P/N distinction. Since the two agents are doxastically and phenomenologically identical to each other, it cannot be either the agent’s phenomenology or the agent’s other beliefs that determine which beliefs are privileged for that agent and which beliefs are not.

There are a number of introspectible features that typically accompany our perceptual, but not our scientific or philosophical beliefs, for example, and it might be
tempting to try to spell out privilege in terms of some of these features. Perception (for us, at least) involves being appeared to in some way; something's seeming red to me is phenomenologically rich in a way that nominalism's seeming true is not. Furthermore, most of us have (at least tacit) beliefs about perceptual beliefs, including, perhaps, that they do not require a generally adequate argument. These suggestions fail to provide an account of privileged belief, however, for Norman and Tru Temple are, by hypothesis, doxastically and phenomenologically identical to their alien counterparts. Whatever phenomenology accompanies Nyrmoon's clairvoyant belief also accompanies Norman's clairvoyant belief, and therefore, the phenomenology cannot serve to distinguish Nyrmoon's privileged belief from Norman's nonprivileged belief. And since the two agents are doxastically identical, an analogous argument shows that the agent's other beliefs cannot serve to determine which beliefs are privileged for that agent and which are nonprivileged. Obviously, the same considerations apply to Tru Temple and Viper Temp.

A more general point holds. Since they are doxastically and phenomenologically identical, there would be no introspectible difference between Norman and Nyrmoon or between Tru Temple and Viper Temp. Because of this, and because Norman's belief is nonprivileged while Nyrmoon's belief is privileged, the difference between privileged and nonprivileged beliefs must lie in some non-introspectible difference. These cases, therefore, argue for a kind of externalism concerning the P/N distinction. An epistemic agent cannot tell from the inside whether he is Norman or Nyrmoon, Tru Temple or Viper Temp (since the beliefs and the phenomenology would be the same in either case). Because of this, an agent cannot tell from the inside whether her beliefs are privileged or
So whatever it is that distinguishes privileged from nonprivileged beliefs, it must be something that is not (entirely) introspectible, something that the agent cannot (always) detect from the inside.

1.2 Perception vs. Reasoning; Norman vs. Nyrmoon

So what does distinguish perceptual beliefs, memory beliefs, Nyrmoon’s belief, and Vipertemp’s belief, on the one hand, from our scientific beliefs, our philosophical beliefs, Norman’s belief, and Truetemp’s belief, on the other hand? The only noteworthy difference between the cases has to do with the causal history of the belief and, in particular, with the nature of the process that produced the belief. Although I think that these cases argue for a kind of externalism and for an important epistemic role to be played by processes, I am not trying to move in the direction of a process reliabilist account of the P/N distinction. For we can easily suppose that the process responsible for Norman’s belief is no more and no less reliable than the process responsible for Nyrmoon’s belief, and we will still think that Nyrmoon’s belief is privileged and Norman’s belief is not.

However, when we start with the original Norman case and alter it so that the agent’s belief is now the result of a faculty similar to our perceptual faculties, our intuitions about privilege change. When we hear the original Norman case, we assume that Norman is a human. There are a lot of things we (think we) know about normal humans: that we do not have clairvoyance faculties, that we do have some evidence for the reliability of our

---

2 Certainly, not every version of internalism claims that we can determine on the basis of mere reflection whether or not any given belief has the relevant epistemic property, although some do (at least with respect to justification [see, e.g., Chisholm, 1989]). This is why I say it argues for “a kind of externalism”. I do not intend to say anything very exhaustive or informative about externalism in this chapter (though I will say more in Chapter 6).
perceptual faculties, and so forth. The assumption that Norman is a human invites us to assume that whatever the source of his clairvoyance belief, this belief does not result from a psychological faculty which is at all similar to our perceptual faculties. Perhaps his belief is not the result of a faculty at all (this seems to be the way we pretheoretically understand, e.g., wishful thinking -- another instance of nonprivileged belief). Or perhaps it is the result of some psychological faculty, but one which just arose recently: that would explain why Norman has no evidence for his being a reliable clairvoyant.

The original Norman case is somewhat underspecified, but it seems that something like this is going on, for when we stipulate that the agent in question has an innate clairvoyance faculty which resembles our perceptual faculties, the case no longer serves as a compelling counterexample to simple reliabilism. I suspect that it is because of this underspecification that intuitions about Norman tend to be less robust than intuitions about Truetemp. Although the Tempucomp (the device that produces Truetemp’s beliefs about the ambient temperature) might be plausibly construed as a faculty (of an odd sort), the case explicitly states that Truetemp only acquired this faculty quite recently.

These cases suggest that what determines our intuitions about whether or not a particular belief is privileged is the assumed nature of the psychological faculty that produced the belief. Some faculties (e.g., perception, memory, and Vipertemp’s temperature-sensing faculty) produce privileged beliefs, while other faculties (e.g., reasoning and Truetemp’s temperature-sensing faculty) produce nonprivileged beliefs. We have a pretty good pretheoretical grasp on what faculties we think there are and which of these faculties yield privileged beliefs (at least for humans). Furthermore, we know that none of these faculties produces beliefs like ‘nominalism is true’ or ‘the mass of the top quark is 175 GeV’. Only faculties that are sufficiently (and in the relevant respects) similar
to our perceptual faculties produce privileged beliefs. And it is in this sense that the present account of the P/N distinction is a faculty-oriented account. Whether a belief is privileged or not depends on the faculty that produced (or sustains) that belief.3

Since I am making privilege a matter of the faculty that produced the belief, the present view embodies an etiological account of the P/N distinction. But I want to defend a doubly etiological account, for I think that it is not just the causal history of the belief that determines privilege, but the history of the faculty as well. If we understand Norman's belief as being the product of a psychological faculty at all, then we assume that the faculty is not one that Norman always possessed, although Nyrmoon did always possess his clairvoyance faculty. Similarly, the only apparent relevant difference between Truetemp and Viperemp is that Viperemp has always had a temperature-sensing faculty, and Truetemp has not. Thus, not only are the synchronic properties of the faculty important, but its history matters as well.

There is one interesting difference here that might initially seem to obviate an etiological account, and therefore appeal to the internalist, although in the end it cuts the wrong way. Even though we might suppose that Norman and Nyrmoon, for example, are doxastically identical, the differences in the cases suggests that the two lack beliefs about the reliability of their clairvoyance abilities for different reasons. In particular, there is no way that Norman could know that he has a reliable clairvoyant power (assuming the ability

3 This meshes nicely with a point about perceptual beliefs made at the beginning of Chapter 2. That a particular belief (e.g., 'there's something red in front of me') is a perceptual belief is not something we can always determine from the content of the belief, although it is sometimes possible to tell that a particular belief (e.g., 'nominalism is true') is not a perceptual belief from the content alone. A generalization of this point seems correct: it is impossible to know from the content alone that a particular belief is privileged, although it is sometimes possible to tell from the content that it is nonprivileged. It is a point in favor of the etiological, faculty-oriented account that it explains this asymmetry, both in the case of privilege and a fortiori in the case of perception.
just arose overnight), but the only reason Nyrmoon does not believe he has a clairvoyant faculty is that he is extremely unreflective. However, this difference cannot account for our intuitions, for these facts ought to make Norman’s beliefs more justified than Nyrmoon’s. If this difference between Norman and Nyrmoon had any effect on our intuitions, it ought to reverse them, since Norman, unlike Nyrmoon, seems to be suffering from a case of inculpable ignorance. At the very least, Norman’s ignorance is less culpable than Nyrmoon’s.

We have been supposing that Truetemp and Vipertemp are doxastically and phenomenologically identical, even though the former’s belief is nonprivileged and the latter’s is privileged. The etiological constraint on faculties appears even more clearly if we add the further supposition that Truetemp and Vipertemp are not just doxastically and phenomenologically identical, but are also molecularly identical. This supposition, I think, does not change the intuition that Truetemp’s belief is nonprivileged even though Vipertemp’s belief is privileged (surely our intuition that Truetemp’s belief was unjustified did not hinge on the assumption that the Tempucomp was made of inorganic material). Even if the two are molecule-for-molecule duplicates, the Truetemp case would serve as a counterexample to simple reliabilism while the Vipertemp case would not.

There are only two kinds of features that could differ across molecularly identical agents: diachronic features and environmental features. To see that it is not environmental features that distinguish Truetemp and Vipertemp, we need only to stipulate that the (synchronic) environments are identical. We can easily suppose that Truetemp and Vipertemp occupy qualitatively identical physical environments, despite their different

---

4 This response is really a version of the ex ante undermining approach that I considered and rejected in Chapter 3 (Section 2.1). For the present purposes, however, I will waive the objections offered there.
histories, and this does not change our intuitions about the two cases.

A slightly more difficult problem is their respective social environments. Even though the two are molecularly identical, they are members of very different social groups. Truetemp is the only human to have that kind of temperature-sensing capacity, while all of Vipertemp's conspecifics have it. To show that privilege cannot be determined by the agent's social group, we need merely to notice that there could easily be agents who have no social group but still have privileged and nonprivileged beliefs. Even if Truetemp had spent his entire life as the only sentient creature on his planet (and the Tempucomp arose spontaneously, or was implanted by aliens, without his knowledge), we would think that his perceptual beliefs are privileged but that his temperature beliefs are not.

1.3 The Generic Etiological, Faculty-Oriented Account

If it is neither the agent's physical environment nor her social environment that determines which of her beliefs are privileged and which are nonprivileged, this seems to leave only the agent's history -- and in particular, the history of the relevant faculty -- that can distinguish between molecular duplicates. This leaves us with a family of etiological, faculty-oriented accounts of the P/N distinction. Whether or not a belief is privileged is determined by the nature of the faculty that produced that belief:

(EFA): A belief is privileged if it is the output of a faculty that (a) is similar to our perceptual faculties in the relevant respects and (b) has the right etiology; all other beliefs are nonprivileged.

EFA leaves quite a bit unspecified. I have not (yet) said what exactly a faculty is, what the right etiology is, or what the relevant respects are in which faculties that produce privileged beliefs must resemble our perceptual faculties. For the moment, I am
intentionally leaving these out, since the main concern in this chapter is to defend a **generic**
etiological, faculty-oriented account.\(^5\) Despite the absence of crucial detail, the framework
account just offered is surely controversial enough to be nontrivial. Doxastic and
phenomenological accounts of the P/N distinction, for example, are inconsistent with the
etiological, faculty-oriented account, even as sketchy as the latter is. Nonetheless, I have
argued that the cases are better accounted for by the present account than by either of these.

Again, my main concern here is with the generic descriptive epistemological
account offered above. However, I think that a more compelling argument for this generic
account can be given by spelling out and defending a specific instantiation of it. Although I
am more confident in the generic account than in the specific account offered below, I think
that the specific account is at least plausible enough to illustrate the promise of the more
general etiological, faculty-oriented approach.

So far, I have not made much of the fact that the goal here is a **descriptive**
epistemological theory. This will become much more explicit when I defend the specific
account. After this account is developed I will offer a **normative** epistemological account,
partly by way of contrast, and partly because it is important in its own right. Both the
descriptive and the normative accounts below illustrate the fruitfulness of the generic
etiological, faculty-oriented account (EFA), even though the normative account will drop
some of the essential features of the descriptive account.

2. Descriptive Epistemology and OE Faculties

The defining goal of descriptive epistemology is “to elucidate commonsense

---

\(^5\) This generic account functions as a sort of schema that yields a family of distinct
but related accounts as we plug in different details concerning the nature of faculties, right
etiology, and relevant similarities.
epistemic concepts and principles” (Goldman, 1992a; p. 155), thereby systematizing our intuitions concerning knowledge and/or justification. My concern here is significantly narrower than this; I will leave questions of knowledge completely to one side, and I will likewise ignore many questions about our judgments concerning justification in general. The target here is an account of our judgments concerning privileged and nonprivileged beliefs. Following Goldman (1992a), once we clearly distinguish this purely descriptive project from other possible epistemological concepts, what emerges is a project which seeks to explain and predict our intuitive judgments concerning justifiedness (or in this case, concerning privileged or nonprivileged status) by appealing to the “knowledge structures” that are operative in producing these intuitive judgments. The capacity to form judgments about epistemic properties is clearly just a special case of a more general capacity to categorize things, and for this reason, Goldman sees the concept literature in cognitive science as having some potential bearing on this project (1992a, p. 156). If psychology can offer any insights into the structure of our concepts or the nature of our classifying capacities, then the descriptive epistemologist should be responsive to these findings.

2.1 Conservatism and Intuitive Judgments

Elaborating on Goldman’s proposal, I want to point out one particular kind of insight that cognitive science offers (though there is also plenty of anecdotal evidence for this). Our intuitive judgments -- and here it is imperative to distinguish between intuitive judgments and judgments that come as the result of deliberation -- seem to be more a kind of “knowing how” than “knowing that”; that is, they seem to rely on a kind of procedural, 

---

6 See Chapter 3, p. 101, for some caveats about the use of the term ‘knowledge structures’.
rather than declarative knowledge. This procedural knowledge seems to be quite uninfluenced by many of our explicit, occurrent beliefs. Subjects in psychological experiments often produce intuitive judgments which they simply would not have made, had all of their beliefs been operative in producing the judgment.

Take two interesting examples, one probably more familiar to epistemologists than the other. In certain applied tasks involving either deductive reasoning or probabilistic reasoning, informed subjects, with some training in logic and the probability calculus, do little or no better than naive subjects (see, e.g., Nisbett and Ross, 1980, for an older but still relevant overview of this literature). Knowledge of the rules of deduction or of the theorems of the probability calculus seems to have little effect on these intuitive judgments. Another example involves studies of what is often called either “intuitive physics” or “naive physics”. We are capable of forming judgments about the trajectories of moving objects, for example, even though many of these intuitive judgments are in direct conflict with the most established laws of physics. Interestingly, high school physics teachers, who were clearly aware of the laws of linear and angular momentum, performed little or no better than subjects with no formal training in physics (Proffitt, Kaiser, & Whelan; 1990). Again, it seems that knowledge of the relevant laws of physics has little bearing on the intuitive judgments themselves.

I will assume without argument that our ability to produce these intuitive judgments

---

7 As in the discussion of knowledge structures, ‘knowledge’ in this context is not intended to have any epistemic connotations.

8 For example, Tversky & Kahneman (1983) found that the conjunction fallacy was committed as frequently among doctoral students with extensive backgrounds in statistics and probability as among undergraduates with no formal training. Cheng et al. (1986) found that subjects who had completed an introductory deductive logic course failed to perform significantly better on the Wason card selection task than subjects who had not taken the course.
(however inaccurate these judgments may be) results, at least in the typical case, from the existence of some knowledge structures. What the above examples tend to show is that agents often have in their possession additional knowledge structures (e.g., information about the rules of probability or about the laws of physics) which, for some reason, are not operative in producing the intuitive judgments (though they very well might be operative in producing deliberative judgments). Our intuitive judgments, whether they are about physics or probabilities, are often insensitive to relevant information that we clearly possess. The fact that $S$ sincerely believes that $L$ is a law of physics is consistent with the fact that $S$ also possesses a knowledge structure which both (a) is responsible for her judgments concerning the trajectories of objects and (b) contains some principle $L^*$ which is obviously inconsistent with $L$. This is a robust finding in psychology; across a large number of domains, we see that intuitive judgments rely on knowledge structures that are resistant to change in the face of new information.

What does this have to do with descriptive epistemology? Recall Goldman's claim, mentioned in Chapter 3, that the knowledge structures involved in our intuitive judgments of justifiedness are not susceptible to capricious change. We can now see that this move is far less ad hoc than it might have first appeared. Perhaps the knowledge structures responsible for our intuitive judgments of justifiedness are susceptible to emendation, but we should expect them to be resistant to it. In particular, we cannot conclude from the fact that attributor $A$ believes that $p$, that this belief will have any effect whatsoever on $A$'s intuitive judgments. And if intuitive epistemological judgments are anything like intuitive probabilistic or physics judgments, the typical attributor will have no idea what information

---

9 For most cognitive tasks, there is simply no alternative explanation available for how we manage to do what we do.
is embedded in the knowledge structures that are responsible for these judgments.

This is an essential fact to keep in mind as we engage in descriptive epistemology, and not only for the use to which Goldman wants to put it. In what follows, I will attempt to capture our intuitive judgments by appealing to knowledge structures which contain principles that neither I nor, I suspect, much of my audience explicitly, consciously believes. It is imperative to remember that even though an attributor wholeheartedly believes that $p$ is false, she very well might still form intuitive judgments that proceed from knowledge structures that contain $p$. In saying this, I do not mean to imply that the relevant knowledge structures are completely unresponsive to any of the agent’s other beliefs; it is just that these knowledge structures appear to change, if at all, then only very slowly and in response to beliefs that are both longstanding and deeply held.

I have promised that the descriptive epistemological account of the P/N distinction would have something important to do with cognitive faculties. The notion of cognitive faculties, therefore, needs some explication. Since, however, it is the descriptive epistemological project that is at issue in this section, the following is not at all intended as a characterization of what these things really are. The goal, roughly, is to elucidate some deeply entrenched beliefs about what they are. More accurately, the goal is to explain the belief-like informational states that figure into the knowledge structures responsible for our intuitive judgments about privileged and nonprivileged beliefs. The main criterion for success here is that an account emerges which predicts the intuitions to a tolerable degree of accuracy. It is less important that the principles invoked here strike the reader as having a great deal of independent plausibility, since these principles are not supposed to be the sorts of things to which an attributor has conscious, introspective access. Hence, when I speak of our “pretheoretical” or “commonsense” understanding of this or that, I do not intend this
sort of claim to be assessable through introspection. The principles invoked in this section will, I think, seem familiar, which is what we would expect if the knowledge structures in question developed out of erstwhile conscious, explicit beliefs (which I suspect is at least partially the case), but such familiarity is not necessary for the success of the present account.

Another factor serves to exacerbate this difficulty. I have already argued that knowledge structures encoding information about cognitive faculties are relevant to our intuitive judgments about justifiedness. Now, several authors have claimed that our intuitive judgments about the mental states of others (and perhaps even ourselves) are determined by knowledge structures that encode what amounts to a psychological theory, often called “folk psychology”, or “naive psychology” (e.g., Gopnik, 1993; Gopnik & Meltzoff, 1997; Stich & Nichols, 1992). Even if this is true, there is no guarantee that the knowledge structures involved in forming intuitive judgments about the mental states of others are the same knowledge structures that are involved in determining whether a belief is privileged or not. This is true even if, as I argue, the latter task involves knowledge structures encoding information about the nature of the mind. The psychological principles (about what faculties there are and what properties they have) that enter into our intuitive judgments about justifiedness might be completely separate from the ones that are responsible for our intuitive judgments about the mental states of others. In other words, there might not be just one “folk psychology”.10 Thus, the assumptions about the nature

---

10 Or, there might be just one, but for a very different reason. Defenders of the simulation theory, e.g., Goldman (1992b, 1993); Gordon (1986), might claim that there is a psychological theory of sorts that is involved in our intuitive judgments about privilege, but not in our judgments about the mental states of others. (It is hard to see how the simulation mechanism would account for beliefs about what faculties there are and what properties they have.)
of the mind to be discussed here not only are not entirely assessable via introspection; they are not entirely assessable by appeal to the cognitive scientific literature on naive psychology, without some independent evidence suggesting that the same knowledge structures are involved in both tasks.

2.2 A Commonsense Understanding of Faculties

The term 'faculty' is often used in two importantly different senses, for a cognitive faculty can be either a mental entity or a mental capacity. The difference here is the difference between parts of the mind and mental dispositions. Plato, for example, clearly thought that the mind was divided into faculties qua mental entities or parts of the mind, as did Gall, and as do many contemporary psychologists. Hume and Reid, on the other hand, when talking about faculties, most likely intended such talk merely as a convenient shorthand for describing certain mental phenomena.\(^{11}\) One way to see the difference between these two uses of the term is to note that if one is to explain someone's perceptual abilities by appeal to their having certain perceptual faculties, one had better mean "faculties" in the former sense, for the latter sense just is that of perceptual abilities. On a dispositional understanding of faculties, such an explanation would be like Moliere's famous example of explaining the sleep-inducing powers of morphine by appeal to its dormitival virtue. If we understand faculties as mental entities, on the other hand, claiming that S has a perceptual faculty does, at least to some extent, explain S's ability to have perceptual states. Henceforth, when I use the word 'faculty', I will mean in the sense of a

\(^{11}\) Sosa (1991b) seems to use the term 'faculty' in this dispositional sense, since he thinks that there is an "infallible cogito faculty" (p. 138) for grasping one's own current existence.
mental entity, rather than a mere disposition.\textsuperscript{12}

This distinction makes it possible to articulate a view, which, though almost certainly false (or at least radically incomplete), is deeply ingrained in our commonsense understanding of the mind. Our commonsense conception of the mind sees it as containing a number of faculties (qua mental entities): memory, imagination, vision, audition, reason, volition, and so forth.\textsuperscript{13} These faculties we take to be parts of the mind, not merely things the mind can do.

An important feature of our perceptual and memory faculties is that they produce beliefs that are what BonJour (1985) calls “cognitively spontaneous”. “I do not infer that there is a red book on the desk, nor does the belief result from any other sort of deliberative or ratiocinative process, whether explicit or implicit. Rather it simply occurs to me, ‘strikes me,’ in a manner which is both involuntary and quite coercive” (p. 117). The term ‘cognitively spontaneous’ is potentially misleading, for paradigmatic cases of cognitively spontaneous beliefs, e.g., perceptual beliefs, are anything but spontaneous in the sense of random and without any particular cause. In fact, it is part of our commonsense understanding of the mind that perceptual beliefs, unlike beliefs that simply “pop into one’s head”, are the result of the operation of a psychological faculty. Nor does cognitive spontaneity require that the belief occur without the accompaniment of any other mental states. The cognitively spontaneous beliefs that are produced by our perceptual faculties

\textsuperscript{12} This distinction between faculties qua mental entities and faculties qua mere dispositions is not the same as Fodor’s (1983) distinction between horizontal and vertical faculties. Although the question of whether there are vertical faculties may become extremely uninteresting on a dispositional reading of ‘faculties’, talk about horizontal faculties can be understood as either talk about mental entities or as talk about dispositions.

\textsuperscript{13} The reason I claim that this view is almost certainly false is not that it distinguishes a number of distinct mental entities, but that it draws the lines in the wrong places.
are accompanied by -- perhaps even preceded and caused by -- sensory, or experiential, states. Still, these perceptual beliefs are not the consequence of an introspectible train of reasoning from these states.\textsuperscript{14}

I will use the terms ‘ratiocinatively opaque cognitive faculty’ or simply ‘opaque faculty’ to refer to any faculty, all or most of whose doxastic outputs are cognitively spontaneous beliefs, i.e., beliefs that are not introspectibly the result of reasoning processes.\textsuperscript{15} When I say of a faculty that it is non-opaque, I do not mean to imply that all of its inner workings are introspectively available, nor do I mean to imply that such a faculty never produces cognitively spontaneous beliefs; I only mean that it is not an opaque faculty. A non-opaque faculty is simply one for which it is \textit{not} the case that all or most of its doxastic outputs are cognitively spontaneous beliefs. Our reasoning faculty is, at least according to our pretheoretical understanding of the mind, non-opaque in this sense, even though there may be a great deal going on inside the faculty that we are not aware of and even though the faculty might sometimes produce cognitively spontaneous beliefs.

\textsuperscript{14} I have to confess that I do not know quite what BonJour means when he claims (in the above quotation) that perceptual beliefs are not the result of even an “implicit” ratiocinative process. If by ‘implicit’ he means “unconscious”, then I will be relying on a slightly different use of his term, “cognitively spontaneous”, for I suspect that our perceptual beliefs are the result of unconscious reasoning, and I would not want to build any empirical assumptions to the contrary into the very notion of cognitive spontaneity. What is important about our perceptual beliefs is that this reasoning is completely inaccessible through introspection, and that is all I will require of cognitive spontaneity.

\textsuperscript{15} The ‘or most’ qualification here will generally not matter, since it seems that, as a matter of fact, all of the beliefs produced by our actual perceptual and memory faculties are cognitively spontaneous. I include the qualification only by way of anticipating possible cases involving a faculty just like a normal human’s perceptual faculty except that every once in a great while, it produces a belief that is not cognitively spontaneous. I would want such a faculty to be classified with a normal perceptual faculty. It is worth pointing out that the notion of opacity will thus be one that might come in degrees. So too might the notion of privilege, although I will ignore these potential complications here.
Our commonsense view of the mind also includes certain expectations about the origin of faculties. We think of faculties in real life as coming about as the result of learning or innateness, although in certain abnormal or fictional circumstances, it is possible to acquire a faculty as the result of surgery or pathology. I will call the expected etiologies “normal” etiologies: a faculty with a normal etiology is one that arises as the result of learning, innateness, or some interaction of the two. The possibility of developing a faculty through learning seems fairly unproblematic. Suppose that, through long years of intensive training, certain Eastern mystics were able to develop a clairvoyance faculty. There seems to be nothing incoherent about such a supposition. The notion of innateness at play here is a bit less straightforward. Although I doubt that this is the correct way to understand innateness -- again, this is descriptive epistemology -- I think our pretheoretical understanding of innateness is, roughly, of a trait’s existing since (around the time of) its possessor’s birth. This understanding of innateness, like the other claims of this section, will receive its primary defense in the following two sections.

2.3 The OE Faculty Account

One of the things that our memory beliefs and perceptual beliefs have in common with each other and with Nyrmoom’s clairvoyant belief is that they are the results of the operation of ratiocinatively opaque cognitive faculties. Being the output of an opaque

16 I do not mean for this to embody any very strong commitments about what counts as a perceptual belief or a memory belief. In particular, I am not (yet) suggesting that this is part of what it is to be a perceptual belief (although I will in Chapter 6). I am only assuming that when we are told that a particular belief is a perceptual belief, we assume that the belief is the output of an opaque faculty. This assumption only requires that the paradigmatic instances of perceptual beliefs are instances where the belief in question is the output of an opaque perceptual faculty. Whether it makes sense to claim that S has a perceptual belief which was not the result of an opaque perceptual faculty has little bearing on the present issue.
faculty, therefore, seems to have something to do with the distinction between privileged and nonprivileged beliefs. In fact, the former is nearly a sufficient condition for the latter. When we hear of a case where an agent has a belief like 'it will snow a lot this winter', we conclude that this belief is not the output of an opaque faculty, since there are no opaque faculties (at least not in normal humans) that produce such beliefs. I suggest that this is at least partly why we intuitively think that the belief is nonprivileged. Beliefs that simply pop into one's head are also nonprivileged, and they are not the output of opaque faculties either. We can further distinguish the faculty of reason from the faculty of reasoning. The former, according to our commonsense understanding, is an opaque faculty which delivers intuitions (in the philosophical sense) and beliefs in axioms, while the latter is a non-opaque faculty which takes us from premises to conclusions.

This, again, does not mean that none of the beliefs produced by the faculty of reasoning are cognitively spontaneous. Suppose that Sherlock Holmes reasons to the conclusion that the butler did it, but he reasons so quickly that the belief does not, introspectively, appear to be the result of a reasoning process and is therefore cognitively spontaneous.17 Clearly such a belief is nonprivileged; Holmes will need a generally adequate argument in order for this belief to be justified. Even though this particular belief is cognitively spontaneous, the reasoning faculty which produced the belief is non-opaque. It is not cognitively spontaneous beliefs per se which are privileged (after all, Norman's and Truetemp's beliefs are cognitively spontaneous); it is beliefs produced by opaque faculties (and only some of these) which are privileged.

As mentioned earlier, the claim that the beliefs produced by opaque faculties are privileged is actually too strong. For suppose that Norman wakes up one day not just with

17 This case was suggested by Alvin Goldman.
a clairvoyance ability, but an actual faculty -- an opaque one -- although he has never before had either the faculty or the ability. Clearly his clairvoyance beliefs are not privileged. The Norman case (or at least this particular version of it) illustrates that not all beliefs produced by opaque faculties are privileged. However, the Nyrmoon case, and the case of Truetemp’s alien counterpart, suggest that the outputs of innate opaque faculties are privileged, since the only feature that distinguishes these agents from Norman and Truetemp is the etiology of their respective faculties.

The notion of innate, opaque faculties will sound familiar to readers familiar with Fodor’s (1983) discussion of modules, since a Fodorian module is simply a particular kind of innate, opaque faculty. This particular etiological constraint, however, can be relaxed somewhat. If the mystics mentioned above were able to develop a clairvoyance faculty as the result of learning, the beliefs produced by such a faculty would intuitively be on a par with Nyrmoon’s clairvoyant beliefs, even though this new faculty is clearly not innate. What the mystics have in common with Nyrmoon is that the relevant faculties have normal etiologies, where, again, a normal etiology includes both learning and innate developmental processes. Thus, I will use the term ‘OE faculty’ to refer to a ratiocinatively opaque cognitive faculty that has a “normal” etiology, i.e., the faculty is the result of learning, innate developmental processes, or some combination of the two. An initially plausible account of the P/N distinction, then, is as follows:

(OEF): A belief is privileged for S if and only if it is the output of one of S’s OE faculties.

OEF, of course, is formally inadequate, since the project under consideration here

---

18 Fodorian modularity will receive more extended treatment in Chapter 5.
is explicitly descriptive. It is easy enough, however, to derive from this a more explicitly descriptive formulation:

\[(\text{OEF}_D): \text{An attributor, } A, \text{ will judge } S's \text{ belief that } p \text{ at time } t \text{ to be privileged with a level of confidence proportionate to the confidence of } A's \text{ judgment that } S's \text{ belief that } p \text{ at } t \text{ is the output of one of } S's \text{ OE faculties.}\]

Even this will require a bit of tinkering, as will be seen below, but I think this is on the right track. For the sake of simplicity, I will take these two to be interchangeable, except when the account actually needs to be cashed out in explicitly descriptivist language.

The main reason for thinking that something like OEF is true is that it gets the cases right. Norman’s and Trutemp’s (nonprivileged) beliefs are not the outputs of OE faculties, because their faculties (if the sources of their beliefs even count as faculties) do not have normal etiologies. Sherlock Holmes’ unconscious reasoning is nonprivileged because it is not the output of an opaque faculty, even though the belief in question might be cognitively spontaneous. Our perceptual and memory beliefs are (or at least so we think) the outputs of opaque faculties (the beliefs they produce are always cognitively spontaneous) with normal etiologies. If we are told that a belief is held on the basis of wishful thinking, we will judge the belief not to be the output of an OE faculty. We (as attributors) might not have a very good sense of what faculty is responsible for wishful thinking (such beliefs might be deemed to be the product of the non-opaque faculty of imagination, or they might not be thought to be the product of any particular faculty at all), but we do have some idea of what OE faculties there are, and no belief produced by any one of them would count as wishful thinking.

\[\text{OEF is also liable to yield intuitively incorrect answers, for reasons to be discussed in Section 3.1.}\]
A somewhat more general argument is gotten by expanding on a discussion from Chapter 2. Recall that a pure perceptual belief is one for which the agent has no independent (non-perceptual) evidence. Similarly, we can define a notion of pure OE faculty beliefs: a belief is a pure OE faculty belief for \( S \) if and only if it is the output of one of \( S \)'s OE faculties and \( S \) has no independent evidence for that belief. Now if there are (many) justified pure OE faculty beliefs, they must be privileged beliefs, for the same reason that if there are justified pure perceptual beliefs, then they would have to be privileged.

If I lack a generally adequate argument before the belief occurs to me, then I will still lack a generally adequate argument after the belief occurs to me, since the belief in question is cognitively spontaneous. Cognitively spontaneous beliefs are beliefs that come to awareness without an introspectible train of reasoning, and therefore, at least typically, without a generally adequate argument. Now there might be an occasional pure OE faculty belief that comes to awareness along with a generally adequate argument (either because it is one of those rare cases of non-cognitively spontaneous beliefs which the notion of opacity allows for, or because it just so happens that at the same time as the belief occurs to the agent, a generally adequate argument occurs to her as well), but it is clear that the vast majority of pure OE faculty beliefs will not be like this. They will come to consciousness with no evidence except their own existence or perhaps a corresponding experiential state. And as we saw in Chapter 2, the inference from the existence of the belief to the truth of the belief, or from the experiential state to the truth of the corresponding belief, is not generally adequate.

For this reason, if there are more than a few justified pure OE faculty beliefs, then they must be privileged beliefs. This argument does not show that all pure OE faculty
beliefs are privileged -- only that the justified ones are. But this is enough to show that skeptical consequences will follow from any view which denies that pure OE faculty beliefs are privileged. The scope and ensuing unpalatability of this skepticism will, of course, depend on how many beliefs we think are the outputs of OE systems. This sort of argument will become more important later, when the topic shifts from descriptive epistemology to normative epistemology.

Before turning to substantive objections, I want to consider an initial objection which shows that OEF_D, as stated, is incorrect. Suppose that you think that belief in God is nonprivileged; that is, you think that no one is epistemically justified in believing in God without some kind of generally adequate argument. (This, I expect, is a pretty standard view.) Suppose, however, that you then hear that a group of cognitive neuroscientists claim to have discovered a “God module” (Ramachandran, et al., 1998). Suppose that they claim that there is an opaque faculty in normal humans that produces belief in God, and suppose that you believe this empirical claim. If this opaque faculty exists in normal humans, then it must have the right etiology and thus must be an OE faculty. If OEF_D is right, then you should suddenly be willing to treat at least some beliefs in God (the ones that are the result of this faculty, though perhaps not the ones that result from reading Aquinas) as privileged. This seems unlikely.

The problem here is that OEF_D, as formulated above, does not specify narrowly enough the relevant doxastic states of the attributor. As mentioned in Section 2.1, our intuitive judgments about justification presumably result from some knowledge structures which are relatively insensitive to change. It is not A’s beliefs per se that determine whether

---

20 Ramachandran and his colleagues actually make a more modest claim than this; I am elaborating on their claim here for the sake of illustration.
A will take a particular belief to be privileged or not, but only those belief-like informational states that are embedded in the knowledge structures that produce A's intuitive judgments about privileged status. Since the knowledge structures involved in these intuitive judgments are expected to be resistant to change, we should not expect this new piece of empirical data to alter our intuitions.

This raises some tricky issues. Obviously some of the attributor's conscious, explicit beliefs have an effect on her intuitions about justifiedness. If nothing else, the explicit beliefs that encode the description of the case (e.g., that Truemp has a Tempucomp implanted in his head) must serve as inputs to the knowledge structures that produce the intuitive judgments about that case. The whole point of distinguishing between what the attributor explicitly believes from what belief-like informational states are embedded in the relevant knowledge structures, however, is that not all of the agent's beliefs are fed in as input. This is part of the lesson of the cognitive psychological experiments discussed in Section 2.1 above.

The beliefs that are fed in as input might conflict, to varying degrees, with the informational states embedded in the relevant knowledge structures. The present account therefore predicts a continuum of intuitions, based on the extent to which the assumptions built into the description of the case conflict with the assumptions built into the relevant knowledge structures. There should not be any conflict in the Nyrmoon case, for example, since presumably the relevant knowledge structures do not contain any information about the nature of alien minds (and certainly nothing about Nyrmoon). Being explicitly told in the description of the case that some particular (unusual) human possesses a clairvoyance module or a God module ought to produce an intuition somewhere between the intuitions we have for the (original) Norman case and the Nyrmoon case, since we pretheoretically
assume that normal humans do not have God modules or clairvoyance modules. Such information, left out of the case description, but recently adopted by the attributor, should have little or no effect on her intuitions.

OEF\(_D\) should be reformulated to take such considerations into account:

\[ (OEF_{D2}): \text{An attributor, } A, \text{ will judge } S'\text{'s belief that } p \text{ at time } t \text{ to be privileged to the extent that the knowledge structures operative in the production of } A'\text{'s intuitive judgments about justifiedness credit } S'\text{'s belief that } p \text{ at } t \text{ with being the output of one of } S'\text{'s OE faculties.} \]

Although this is an improvement over OEF and OEF\(_D\), it is a bit awkward. Consequently, in what follows, I will ignore these complexities and focus on the more general OEF strategy.

2.4 Objections

I think that the general strategy embodied by OEF is right, since OEF latches onto the only obvious difference between Norman, Truetemp, and similar cases on the one hand, and the alien counterparts and normal memory and perceptual beliefs on the other hand. The above cases suggest that whether a belief is privileged or not has something important to do with the nature of the psychological faculty which produced the belief, and OEF strikes me as a promising way to flesh out an etiological, faculty-oriented account of the P/N distinction. However, I do expect some resistance both to OEF and to the more general etiological, faculty-oriented approach to accounting for the P/N distinction.

Although still contentious, OEF would be considerably less important if it only stated a sufficient condition for a belief's being privileged. A merely sufficient condition would not quite do the work than an account of the P/N distinction is needed to do. In
particular, it would not fully explain the difference between Norman and Nyrmoon, or
between Truetemp and his alien counterpart. This is because although a sufficient
condition would explain why Nyrmoon’s beliefs are privileged, it would not explain why
Norman’s beliefs are nonprivileged and thus would not explain why Norman is not
justified.\textsuperscript{21} Is being the output of an OE faculty really necessary for a belief’s being
privileged?

Some initial counterexamples to OEF as a necessary condition for privileged belief
focus on the etiological constraint. Suppose, for example, that a Swampman (who is my
molecular duplicate) arises spontaneously one day as the result of random convergence of
molecules. Surely the same beliefs would be privileged for him as for me, since he is my
molecular duplicate, even though none of his faculties have the sort of etiology OEF
requires.\textsuperscript{22}

Although I tend to think that the Swampman’s perceptual beliefs and the like are
privileged, I do not want to endorse the principle that if \(x\) and \(y\) are molecular duplicates,
then whatever beliefs are privileged for \(x\) must be privileged for \(y\) and vice versa. In fact, I
argued above, in Section 1.2, that there are cases of molecular duplicates that differ in

\textsuperscript{21} I do take some comfort in noting that this is everyone’s problem and not just
mine. Chisholm (1966, 1977), for example, offers sufficient conditions for a belief’s
being directly justified by an appearance belief, but since he never offers necessary
conditions, his account cannot explain why Norman’s belief that he is appeared to
clairvoyantly in a the-president-is-in-New-York sort of way does not directly justify his
(unjustified) clairvoyant belief that the president is in New York. Nor, as was noted in
Chapter 3, can coherence theories explain why Truetemp’s belief in his own reliability is
not enough to \textit{prima facie} justify his belief, absent some necessary condition for privilege.
Probably because the P/N distinction is so much less explicit in coherence theories, no
coherece theory that I know of even attempts to offer a \textit{sufficient} condition for a belief’s
being privileged.

\textsuperscript{22} This particular example is due to John Pollock (in conversation), though the
more general idea of the Swampman comes from Davidson (1986).
epistemic status and that this actually provides further argument for an etiological constraint. After the surgery, Truetemp and Vipertemp have qualitatively identical faculties (since faculties supervene on physical properties and the two are molecularly identical), yet beliefs that are privileged for the alien are nonprivileged for Truetemp.

One notable difference between the Truetemp case and the Swampman case is that Truetemp existed prior to the surgery. To see that this difference actually has an effect on our judgments of justifiedness, notice that if the surgeons who altered Truetemp had instead decided to create an entirely new being out of scratch (call this being “Newtemp”), one which was molecularly identical to Truetemp, I think that our intuitions here would be much as they were for the Swampman. The relevant beliefs do strike us as privileged, although our intuitions here are not quite as strong as they are in cases of genuinely innate faculties.

The reason our intuitions are not quite as strong for the new creatures as they are for normal humans or aliens, I think, is that the faculties in question are innate only in an extended, somewhat unusual sense. Innateness, recall, is pretheoretically understood as presence from the time of birth. The faculties that the Swampman and Newtemp come into the world with count as only marginally innate, because the agents’ coming into existence counts only marginally as having a birth. This would explain why our intuitions about privilege are somewhat less robust for the Swampman and Newtemp cases than for other cases; the relevant beliefs are the outputs of OE faculties, but only barely so and in a somewhat extended sense.²³

²³ The general etiological, faculty-oriented framework leaves room to maneuver here. An alternative treatment of this sort of case might parallel the general structure of Two-Component Reliabilism. One could claim that attributors make categorical judgements about which faculty token produced the target belief and then use a matching procedure to determine whether this token is an OE faculty or not, comparing it to some paradigm OE
At first glance, it might seem that one could use this response to generate additional counterexamples to OEF, for the present response implies that for a being who suddenly comes into existence, all of the opaque faculties she has at the moment she comes into being will be OE faculties. Thus, for such cognizers, any belief that is the output of an opaque system is going to be privileged. This consequence does follow, although I see nothing wrong with it. Suppose that the new creature is psychologically just like a normal human. There is nothing counterintuitive about classifying the outputs of these opaque faculties as privileged. A normal human possesses only faculties with normal etiologies, so as a matter of fact, all of a normal human's opaque faculties are going to turn out to be OE faculties as well. And we have already seen (in the Newtemp case) that the opaque faculties possessed by these new creatures do, intuitively, deliver privileged beliefs, even when such faculties are not shared by normal humans. So either way, there does not seem to be anything counterintuitive about the results.

Another counterexample aimed at the etiological constraint forces a clearer specification of what a faculty is. Suppose that the neural circuits in my visual cortex are replaced, gradually and without my knowledge, by some ingenious neurosurgeons. They could even replace my whole brain with silicon hardware, but then none of my faculties would have the right etiology. Still, some of my beliefs, e.g., my perceptual beliefs, would remain privileged. Thus, the etiological constraint on the faculty that produces the belief does not capture a genuinely necessary condition for privileged belief.24

---

24 This case was also suggested by John Pollock.
I have not yet said anything about the relation between faculties and brains. In the next chapter I will say quite a bit more about this, but for the present purposes it is enough to notice that faculties are not identical to brains or parts of brains. Faculties may be thought of as virtual machines that supervene on brains (at least typically, although they can in principle supervene on other things as well). Thus, certain changes in the underlying hardware might very well leave the original faculty intact. Faculties are multiply realizable, and it seems quite plausible that certain changes in realization are compatible with the identity of the faculty being preserved. Now I do not have an account of the diachronic numerical identity of faculties (or of anything else, for that matter). Still, this strikes me as a plausible, though admittedly sketchy, response. Note that the difference between this scenario and the Truetemp scenario is that in Truetemp's case, some hardware is being added, and this hardware performs a function that the brain did not previously perform. The Truetemp case must be interpreted as the addition not only of new hardware, but of a new faculty as well.

This response may be a bit too sophisticated for the purposes of descriptive epistemology. The folk do not understand virtual machines and multiple realizability (or at least let us suppose). However, all that is needed to forestall the present objection is to note that our commonsense understanding of the mind does not identify faculties with brain parts. According to our commonsense understanding, faculties can arise as the result of learning; brains cannot. As long as faculties are not taken to be identical to brains, it is possible that changes in the brain might leave the relevant faculties intact.

Another class of counterexamples attempts to show that the etiological constraint is not sufficient to make the outputs of opaque faculties privileged. Earlier, I explained the nonprivileged character of wishful thinking by claiming that such beliefs, whatever their
source, are not the outputs of OE faculties, and we believe this because we have a pretty good idea of which OE faculties there are, and if a belief is produced by one of these faculties, then it is ipso facto not based on wishful thinking. But this invites the following counterexample: suppose that some creature did have an opaque wishful thinking faculty with the right etiology. OEF implies that the beliefs produced by that faculty would be privileged, but surely this is wrong.

Now, this might seem a bit perverse at first glance, but I tend to think that when we get clear enough on what is involved in this counterexample, this result is not so implausible after all. The reason OEF seems to yield the wrong answer is that it is hard to get a very good grasp on what a wishful thinking faculty might be. Suppose that we build a creature which is otherwise just like us, except that we put in a special computational module that takes desires as inputs, filters out some of the weaker desires, and produces as output beliefs with the same propositional content as the stronger desires. The claim that the beliefs produced by this module are privileged does not strike me as particularly unpalatable, especially when it is recognized that they might still fail to be justified for a number of reasons. These beliefs might be prima facie but not ultima facie justified, because, for example, the agent has noticed that all of its stronger desires become cognitively spontaneous beliefs and that these beliefs almost always turn out to be false. This would provide undermining evidence that would defeat the initial prima facie justification that accrues to the beliefs as a result of their being outputs of an OE faculty.25

To rule out this possibility, consider some wishful thinking belief that occurs to our creature immediately after it comes online and before it has any undermining or contrary

25 The reliabilist will have a much simpler response here: wishful thinking is not a reliable process, and reliability is necessary for prima facie justification.
evidence. We have built the creature such that a strong, "innate" desire causes a corresponding belief, and the creature has absolutely no evidence against this belief. It would seem unfair to blame the creature for holding this belief, provided, of course, that this belief has whatever other properties are needed to get from being a privileged belief to being a *prima facie* justified belief. What these other properties are will (as mentioned in previous chapters) depend on what more general epistemological view is true -- reliabilists will require some sort of reliability, coherentists will require some sort of otherwise substandard argument, and so forth.

Any causal/historical account of any epistemic property is potentially subject to demon-world counterexamples, since a Cartesian demon could simulate all the internally accessible features of the relevant property, while bypassing the normal causal history. Suppose that a Cartesian demon is responsible for all of my putatively perceptual beliefs. None of these beliefs, or so goes the objection, would be the outputs of OE faculties, yet, clearly, these beliefs remain privileged nonetheless.

Now I think that our intuitions concerning the epistemic status of the demon-produced beliefs depend crucially on how we conceive of the demon causing those beliefs. Compare two extreme versions of the demon scenario: in the one case the demon directly stimulates my sensory surfaces, merely bypassing the external part of the causal chain that normally links objects to perceptual beliefs. Clearly in this case, my belief is privileged, but the belief is just as clearly the output of one of my perceptual faculties.26 Suppose that in the other case, however, the demon simply puts the belief in my head, so that the belief

26 The same holds true if the demon starts a little further down the line: say, implanting in me a perceptual state, which then causes the relevant perceptual belief to occur. The perceptual belief would still be the output of an OE (perceptual) faculty, even though that faculty did not, in this case, receive the normal input.
is in no way caused by the operation of any perceptual faculty. I have no perceptual experience, nor am I subject to sensory stimulation, but nevertheless, the belief that there is something red in front of me just pops into my head from out of nowhere. It is hard to see how the resulting belief would be any different from Truetemp's problematic belief. OEF would classify such beliefs as nonprivileged, and this seems intuitively correct.

There is an intermediate sort of case which I think is somewhat more difficult to handle. Suppose that the demon bypasses the normal processing of visual information, while still making sure each of the normal stages of processing occur. For example, the demon produces optical stimulation on my retina and then interferes with the normal causal chain in such a way that this stimulation now fails to produce a perceptual experience in the regular way. Now the demon gives me the perceptual experience that I would have had, had the normal causal chain not been interrupted. The demon repeats this whole process for the chain from perceptual experience to perceptual belief.27

I have to confess, I do not have a clear intuition about whether this belief is privileged or nonprivileged. Nor, however, do I have a clear intuition as to whether this belief ought to count as an output of a visual faculty. I suspect that our commonsense understanding of what makes something an output is simply not clear on this point. If this seems plausible, and if others also fail to get a clear intuition about privilege from this case, then this would seem to be a point in my favor.

It is tempting in a case like this to reason one's way to an "intuition" in something like the following way:

My demon-world counterpart (in this last case) and I do not differ in any

27 I got the idea for this objection from a similar example described by Haugeland (1980), although Haugeland's case is aimed at something quite different (viz., Searle's claims about intentionality).
introspectible ways; everything is exactly the same for us, from the inside. Since my own belief that there is something red in front of me is privileged, my counterpart's belief must be also.

Such a judgment, however, would no longer be an intuitive judgment but a deliberative judgment, and thus, not the sort of judgment that the descriptive epistemological project is concerned with. (The temptation to come to this deliberative judgment can perhaps be tempered by the reminder that Norman and Nyrmoon are also identical from the inside, despite the fact that Nyrmoon's belief is privileged and Norman's is not.)

Perhaps the most difficult objection to OEF is one that I will call "the no-OE-faculty objection", and it focuses on the opacity requirement. Suppose that there exists a creature that does not have any OE faculties. According to OEF, such a creature would not have any privileged beliefs. This seems too strong. And furthermore, a hypothesis was offered in Chapter 2, according to which all beliefs are either privileged beliefs or ultimately derive their justification from a set of privileged beliefs. If this hypothesis is true, then a creature with no OE faculties would not be able to have any justified beliefs whatsoever. This is definitely too strong.

This is a difficult objection to answer. Obviously, some beliefs would have to be cognitively spontaneous; any introspectible train of reasoning would have to start somewhere. However, OEF requires not just that the belief in question be cognitively spontaneous, but that the faculty be opaque, and it surely seems possible to divide up faculties in such a way that none of them produces all or even mostly cognitively spontaneous beliefs. Suppose there were a creature with only one faculty, and suppose that most of this creature's beliefs were not cognitively spontaneous. This creature would

---

28 This objection is due to Alvin Goldman.
have no opaque faculties, hence no OE faculties, hence, according to OEF, no privileged beliefs. In addition, this objection points to a fundamental vagueness in the notion of faculties. If the notion of OE faculties is going to do the work I want it to, there had better be some fairly principled way of individuating faculties such that perception and reasoning, for example, are two separate faculties; otherwise our perceptual beliefs would not be the outputs of opaque faculties.

Answering this objection is going to require a detailed descriptive epistemological account of faculties. However, I am not yet in a position to offer such an account. My plan is to clarify our ordinary conception of faculties by holding it up against the more precise, cognitive scientific, conception of cognitive systems. The cognitive scientific notion is itself far from precise, but I attempt to make it so in the next chapter. The discussion of faculties will have to wait until then.

Before turning to normative epistemology, let me address one final complaint. Although it is not quite an objection, one might criticize the present view on the grounds that it is either just warmed-over Plantinga (1993a) or derives all its plausibility from tacitly smuggling in a sort of proper functionalism. A proper functionalist could claim that what is really going on is that properly functioning faculties that are designed to function opaquely produce privileged beliefs as their outputs, thus replacing the etiological requirement I favor with a design requirement.

Such a claim would be both too strong and too weak, which is why I have intentionally refrained from endorsing a design requirement. Plantinga's view is subject to counterexamples that mine is not subject to. The Swampman, by hypothesis, has not been designed, and so none of his faculties meet the proper functionalist design requirement (Taylor, 1991). Still, the Swampman does have privileged beliefs, and in this sense, the
design requirement is too strong. Furthermore, there are some faculties that might very well be designed to function opaquely, yet which do not produce privileged beliefs. Suppose, following Lehrer (1996), that Truetemp's surgeons designed the Tempucomp to function opaquely and to produce privileged beliefs. This clearly would not suffice to make Truetemp's temperature beliefs privileged. The problem with the Tempucomp is that it does not have the right etiology, and in this sense the design requirement is too weak, since it does not entail the etiological requirement.

3. Normative Epistemology and Opaque Systems

The normative project in epistemology, recall, is an attempt to transcend -- to improve upon -- our commonsense epistemological principles and concepts. In Chapter 1, I distinguished a means-ends approach to normative epistemology from a reflective equilibrium approach. The means-ends approach is to find out what our epistemic goals are and then find out what is most conducive to them. The reflective equilibrium approach, on the other hand, has us work out an improved epistemology on the basis of our intuitive epistemological convictions. It may not be initially obvious what would count as improvement, if we adopt the reflective equilibrium approach, but I hope that this will become clear later. It is this latter, reflective equilibrium, approach that I will adopt in what follows. I want to use the (admittedly partial) results from the descriptive account just offered and sketch the beginnings of a normative epistemological account of the P/N distinction.

Why adopt the reflective equilibrium approach here rather than the means-ends approach? First of all, it is not clear how a means-ends approach to the distinction between privileged and nonprivileged beliefs might even get off the ground. Since the means-ends
approach is one of endorsing whatever belief-forming methods and practices are most conducive to our epistemic goals, it seems suited only to the formulation of a general epistemological theory of justification. It is hard to see how this sort of approach is going to have anything to say about more restricted epistemological issues, like the nature of the P/N distinction. It is not clear, for example, that we even have any goals associated this.

Secondly, the two authors who are most explicit about distinguishing the descriptive and the normative projects (Goldman, 1992a; Kitcher, 1992) have already utilized the means-ends approach. Kitcher is especially explicit about this, although Goldman, too, seems to endorse a reliabilist normative epistemology as least partly because the use of reliable processes is conducive to the epistemic goal of truth; he even takes the time to argue that truth is an appropriate goal. Goldman does think that the normative account should maintain contact with the descriptive account, but this seems to be more a matter of the normative theory still counting as an epistemological theory than a matter of the normative proposal receiving evidence from the descriptive theory. Thus, another reason for utilizing the reflective equilibrium approach here is simply that it has not been vigorously pursued by contemporary epistemology.

Finally, as I mentioned in Chapter 1, I think that the reflective-equilibrium approach makes better sense of the import of descriptive epistemology than does the means-ends approach. By giving our intuitions about justifiedness a great deal more evidential status than the means-ends approach, it shows us why we should bother with descriptive (and elaborative) epistemology.

3.1 Beyond Folk Psychology

The reflective equilibrium approach attempts both to clarify the relevant concepts
152

and to amend the relevant principles. We take the descriptive epistemological account as a starting point, abandoning principles that seem arbitrary, and clarifying or replacing concepts that seem too vague. A descriptive epistemological theory reveals our deep-seated epistemological convictions as well as other convictions that influence our epistemological intuitions. If the descriptive epistemological account offered above is even in the right ballpark, then it appears that our intuitions are the result of both an epistemological conviction about the relationship between certain cognitive faculties and privileged belief, and a host of non-epistemological (or at least not purely epistemological) unconscious assumptions about the nature of the human mind. Among these are assumptions about what faculties there are and are not as well as assumptions about which faculties are opaque and which are not.

Henceforth, I will refer to this latter set of convictions or assumptions -- the convictions about the nature of the mind -- as our "folk psychology". It should be noted, however, that the term is being used somewhat more narrowly here than is common (see the discussion at the end of Section 2.1). I will use the term to refer only to those beliefs (or belief-like informational states) about the nature of the human mind that are operative in producing our intuitive judgments about justifiedness, and in particular, those intuitive judgments about which beliefs are privileged and which are nonprivileged.

The existence of such psychological convictions suggests an important new way to improve upon the deliverances of the descriptive project: take all the false claims implicit in our folk psychology, and replace them with true ones. If the only reason we intuitively judge some belief to be privileged is that we are mistaken about the nature of the faculty that produced that belief, then this intuitive judgment is one that we should not take very seriously in formulating a normative epistemology. Replacing our false folk psychological
convictions with true psychological facts would provide a clear sense in which the resulting
theory constitutes an *improvement* over the descriptive theory. I want to explore this
possibility here, and to do so, I will assume, for the present, that there is nothing
problematic about the purely epistemological convictions responsible for our intuitive
judgments. (These will be examined in Section 3.2.)

The first step, then, in developing a normative epistemology out of the descriptive
account already sketched, is to return to OEF:

(OEF): A belief is privileged for $S$ if and only if it is the output of one of $S$’s OE faculties.

OEF was formally inadequate for the purposes of a descriptive epistemological theory, for
it said nothing about the attributor. Since the project in question is now normative
epistemology, any such reference to attributors would be inappropriate. Because OEF,
unlike OEF$_D$, omits any reference to the attributor, it makes privilege a matter of the
epistemic agent’s *actual* psychology, rather than a matter of what the attributor *believes* to
be the agent’s psychology. In this way OEF automatically corrects for any ascriptions due
to errors in the attributor’s folk psychology. OEF therefore implies that surprising
discoveries about what OE faculties there are will have surprising results for epistemology.
For example, if it turns out that there *is* a God module in the sense described above, then
OEF will count its outputs as privileged.  

There may be a host of other beliefs that OEF will count as privileged which we
would not have intuitively taken to be privileged. Cognitive scientific research suggests a
number of possible OE faculties, faculties for face recognition (Diamond & Carey, 1986;
Ellis, et al., 1986; Ellis & Young, 1989), cheater detection (Cosmides & Tooby, 1992;

29 This is another reason, alluded to above, that OEF would have been inadequate
for a descriptive epistemological account. Our intuitions are what they are, no matter what
empirical science may discover.
Tooby & Cosmides, 1992), language comprehension (Chomsky, 1988; Forster, 1990),
third-person mental state ascription (Leslie, 1994; Segal, 1996), and many others. It might
even be that our intuitive judgments about probability (i.e., those that rest on the use of
availability and representativeness heuristics [Kahneman & Tversky, 1972; Tversky &
Kahneman, 1973]) are the outputs of such faculties.

It might be that perceptual learning results in more work for our OE perceptual
faculties. Chicken sexing is perhaps the most notorious example of such perceptual
learning. Expert chicken sexers get very good at looking at a chick and determining
whether it is male or female even though they are incapable of articulating the perceptual
features they are using to make this judgment (Gibson, 1969). Although it is clear that any
beliefs I might form concerning the sex of a chick would be both nonprivileged and not the
result of an OE faculty, the judgments of expert chicken sexers might very well be the
output of otherwise normal perceptual faculties. My (OE) perceptual faculties do not
produce beliefs about the sex of chickens, but some peoples’ might. There might, of
course, be other people who are quite good at identifying chicken sex on the basis of an
explicit inference from a perceptual belief. Such beliefs would not be classified as
privileged by OEF, although the beliefs of the expert chicken sexers might be (depending
on whether these beliefs are, in fact, the output of OE faculties).

Other kinds of learning (i.e., nonperceptual learning) might even result in wholly
new (OE) faculties. Experts of various sorts, for example, tend to have cognitively
spontaneous beliefs about things in the domain of their expertise (Karmiloff-Smith, 1986;
Norman & Shallice, 1983; Shiffrin & Schneider, 1977). Such cognitively spontaneous
beliefs are certainly the outputs of some faculty or other, and it is not at all clear that they
are the outputs of perceptual faculties. This possibility gains some plausibility, I think,
when we move from the notion of faculties to the cognitive scientific notion of cognitive systems (which we will do shortly), especially once the latter is made more precise. However, I mention this sort of case only as an interesting possibility and will not discuss it further in any detail.

Now these results do not strike me as particularly counterintuitive. It does not do excessive violence to my intuitions to suppose that beliefs about the identity of faces, the mental states of others, the sex of chickens, and the like might be privileged beliefs. But even if these were to be intuitively unacceptable results, this would not, in itself, suffice for the rejection of OEF as a principle of normative epistemology. Normative epistemology is an attempt to transcend our intuitions, and it could hardly do so if none of its implications were at all counterintuitive. The question, which I will not be in a position to even begin to answer until Section 3.2, is whether these results are worse than abandoning the principles that led to them.

The success of OEF, or something like it, depends on there being a clear answer to the question of whether a particular belief is the result of an OE faculty. This requires a clear specification of what a faculty is, what ratiocinative opacity is, and what the etiological constraint amounts to. For reasons to be discussed shortly, I will have very little to say by way of clarifying the etiological constraint. The notions of faculties and opacity will receive extended treatment in the next chapter. For now, it should suffice to point out that there is a fairly clear scientific notion of faculties at work in cognitive science: the notion of cognitive systems. The notion is worthy of extended treatment, both for the present purposes and because it is of independent interest to cognitive science. Without actually getting into the details of Chapter 5, I can say for now that the resulting notion of cognitive systems offers a principled distinction between various mental entities and not
just various mental capacities or dispositions. The vague reference to faculties can be replaced by reference to the more precise notion of cognitive systems:

(OES): A belief is privileged for S if and only if it is the output of one of S's OE systems, where an "OE system" is simply a cognitive system in the sense of Chapter 5, which is ratiocinatively opaque and which meets the etiological constraint.

### 3.2 Rendering the Epistemological Principles Less Arbitrary

Stich (1990) has argued that many of our commonsense epistemological convictions are idiosyncratic and unprincipled. This, of course, does not spell the death of epistemology; rather, it merely indicates that these convictions should not find their way into a normative epistemological theory. I agree that some of our intuitions are unprincipled. In the last section, I advocated replacing our folk psychology with true psychology (even if the verdict is still out on just what psychology is true). There is no good reason to hold on to our intuitions even in the face of the fact that the only reason we have these particular intuitions is that we are mistaken about the nature of the mind.

Similarly, I think that the purely epistemological principles invoked in the descriptive epistemological account of the P/N distinction are susceptible to revision as well, even if not on the basis of direct conflict with empirical findings. The descriptive account of Section 2 highlighted two important epistemic principles: (i) that some beliefs are privileged and others are nonprivileged, and (ii) that all and only the beliefs produced by OE faculties are privileged.\(^{30}\) Is there anything unprincipled or idiosyncratic about these views?

Let us start with (i). I argued in Chapter two that unless some beliefs are privileged,\(^{30}\) These, of course, are the two main theses promised in Chapter 1.
a kind of skepticism is sure to result, and I argued in Chapter 3 that unless some beliefs are nonprivileged, justification will be too easy. The desire to avoid either extreme seems quite appropriate. There does not seem to be anything intrinsically unprincipled about holding some beliefs to different standards than other beliefs. But perhaps there is something unprincipled about the way we determine which beliefs are which; that is, perhaps there is something unprincipled about (ii).

OES has several components, and each of these should be taken in turn. As an instance of the more general etiological, faculty-oriented account, OES is an etiological theory, and doubly so. First, it makes privilege a matter of the etiology of the belief (i.e., the belief’s being an output of a particular kind of cognitive system). Secondly, it makes privilege a matter of the etiology of the system, due to the etiological constraint on what counts as an OE system. This second etiological constraint does seem unprincipled. Why should the etiology of the system make a difference? I argued in Section 2 that it does make a difference, at least as far as our intuitions are concerned, but the question here is whether we should take such intuitions seriously. I have already argued that our

---

31 The issue here is not one of foreclosing either of these extremes, but, rather, one of leaving room to avoid them. One who wants to endorse skepticism can still do so, within the normative epistemological account I offer, by, e.g., making the requirements for the justification of privileged beliefs extremely high. And one can advocate an epistemological theory of the sort that I would consider too lax, simply by claiming that all privileged beliefs are justified and accompanying this with very low standards for what counts as a generally adequate argument. These extremes can be reached, even within a normative epistemological theory that includes a P/N distinction, but a normative epistemological theory that claims that all beliefs are either privileged or nonprivileged commits one to one of these extremes.

32 Although the descriptive epistemology theory endorsed here is a version of OEF, I have already argued that we should replace the vague notion of faculties with the more precise notion of cognitive systems. For this reason, the following discussion will be in terms of OES.
perceptual beliefs have the same (privileged) status as Vipertemp’s temperature beliefs to be privileged as well. His beliefs are just as cognitively spontaneous as ours, just as reliable as ours, etc. Similarly, although I certainly have the intuition that Vipertemp’s beliefs differ epistemically from Truetemp’s beliefs, I cannot see any reason why this should be so.

Although Truetemp and Vipertemp may be molecularly identical, one existed prior to the appearance of the temperature-sensing system, and the other did not. But I cannot see why this should matter -- why Vipertemp’s beliefs should be privileged while Truetemp’s beliefs are not. Unless there is some principle behind the etiological constraint, any beliefs that are privileged for the one ought to be privileged for the other. My (tentative) response to this is simply to drop the etiological constraint. If Vipertemp’s beliefs ought to be on a par with our perceptual beliefs and Truetemp’s beliefs ought to have the same status as Vipertemp’s beliefs, then Truetemp’s temperature beliefs and our perceptual beliefs stand or fall together. Either they should both be privileged or they should both be nonprivileged. Certainly, it is counterintuitive to claim that Truetemp’s temperature beliefs are privileged, but it would be even more counterintuitive to deny that our perceptual beliefs are.33

The opacity constraint seems different. I think there is a principled reason for maintaining this constraint in a normative epistemology. As argued in Chapter 2, some beliefs have to be privileged if we are to avoid a crippling sort of skepticism. And I argued above, in Section 2.3, that the outputs of opaque systems would have to be privileged in order to be (generally) justified. A normative epistemological theory needs to admit of

33 Again, claiming that Truetemp’s belief is privileged does not imply that it is justified, or even prima facie justified, even though it certainly makes it easier for the belief to be justified.
privileged beliefs or degenerate into skepticism, and I think that the best candidates for privileged status are those beliefs which, due to the nature of the systems that produce them, come to consciousness with no argument except for their own existence and/or perhaps some accompanying experiential state (neither of which provides for a generally adequate argument). These are cognitively spontaneous beliefs. Yet we saw earlier in this chapter that some cognitively spontaneous beliefs might be the result of a non-opaque system. Why deny privileged status to these beliefs? That is, even if skeptical considerations suggest that there is nothing unprincipled about counting some beliefs (even perhaps the outputs of opaque systems) as privileged, why should we require of any privileged belief that it be the output of an opaque system?

The difference between the cognitively spontaneous beliefs produced by opaque systems and those produced by non-opaque systems is that there is some sense in which the latter did not have to be cognitively spontaneous. If Holmes' cognitively spontaneous belief that the butler did it is the output of a non-opaque reasoning system, then Holmes could have simply slowed down, or paid more attention to his reasoning, in order to become aware of the reasoning that led to the conclusion. A cognitively spontaneous belief that results from a non-opaque system is one that, in some sense, did not have to be cognitively spontaneous. But the cognitively spontaneous beliefs that result from the functioning of an opaque system are different. One cannot simply slow down or pay more attention, thereby making one's perceptual beliefs no longer cognitively spontaneous. This difference between opaque and non-opaque faculties seems relevant. We ought to be reluctant to grant privileged status to beliefs, for fear of lowering inferential standards too far, and given this difference between the cognitively spontaneous beliefs produced by opaque and non-opaque systems, this seems like a reasonable place to draw the line.
This suggests that we drop the etiological constraint, but maintain the ratiocinative opacity constraint:

(OS): $S$'s belief that $p$ is privileged if and only if $S$'s belief that $p$ is the output of one of $S$'s ratiocinatively opaque cognitive systems.

All the above argument is intended to show is that there is nothing particularly unprincipled about identifying privileged beliefs with the outputs of opaque cognitive systems. It does not show, and is not intended to show, that this is preferable to any other possible account of the P/N distinction. What the present view has going for it over other possible views is that it maintains contact with the principles that drive our epistemic intuitions. These principles represent our deep-seated epistemological convictions, and using them as a starting point is what distinguishes the reflective equilibrium approach to normative epistemology from the means-ends approach. The reflective equilibrium approach takes our pretheoretical intuitions more seriously than the means-ends approach, assuming, in effect, that the components of the descriptive theory should be maintained unless there is some specific reason not to maintain them.

Normative epistemology is an attempt to improve upon our commonsense epistemological convictions. The above illustrates a fairly clear sense in which the results of reflective equilibrium might constitute an improvement over OEF$_D$2: OS, the result of the reflective equilibrium approach, corrects for errors about the nature of the mind and eliminates epistemic principles that are unprincipled. I do not claim that OS is the result we would get from using the means-ends approach. (As I mentioned earlier, I am not certain whether the means-ends approach would result in any account of the P/N distinction at all.) The above argument does not begin to show that accepting (all and only) the outputs of our opaque systems on the basis of an otherwise substandard argument is especially conducive
to our epistemic goals. But neither was it intended to show this.

4. Reservations

I have offered here two tentative accounts of the P/N distinction: a descriptive epistemological account, and a normative epistemological account. One reason both accounts are as tentative as they are is that too many relevant empirical facts are still unknown. The descriptive account does seem to explain our intuitions, but there might be other, competing accounts that do so as well. Until there is some independent reason to think that we possess the sort of knowledge structures invoked by the present descriptive account, this account has little to recommend it over its possible competitors, except for the fact that there currently are no actual competitors. And the fact that there might be more than one “folk psychology”, as mentioned in Section 2.1, makes it even harder to get this independent evidence.

The normative account also depends on empirical data that are not yet in. Cognitive science is only in the very early stages of being able to tell us what cognitive systems there are and what opaque systems there are. None of the deliverances of cognitive science on this point (that I am aware of) yield completely unacceptable results, but it is certainly not inconceivable that they would. If cognitive science were to discover, for example, that there really is no interesting distinction to be drawn between different systems, or faculties, then the present account would have to be abandoned, since it would imply that all beliefs are nonprivileged.\footnote{This is because if there are not distinct systems within each brain, then each whole brain would realize only one system, and this system would almost certainly not count as ratiocinatively opaque.}

Still, the above strikes me as a reasonable first pass -- if nothing else, a starting
point for future research. Until the relevant empirical findings are in, we can still pursue
the more purely philosophical project of trying to clarify the relevant concepts that enter into
the present account of the P/N distinction. As mentioned earlier, the opaque systems
account developed in 3.2 has a long way to go before it has anything like the sort of
precision we would want from a full-blown epistemological theory. In particular, there is a
lot more to be said about what an "opaque cognitive system" is. This will be the topic of
the next chapter.
CHAPTER 5: FACULTIES AND COGNITIVE SYSTEMS

In Chapter 4, I defended two accounts of the P/N distinction: a descriptive epistemological account, which appealed to cognitive faculties, and a normative epistemological account, which appealed to cognitive systems. The difference between these two notions of faculties and systems is merely that the former notion is involved in our pretheoretical understanding of the sense in which the mind is composed of parts, while the latter is intended to capture a scientific understanding of the same. I have not yet said much about what either a faculty or a cognitive system is, and this is an essential component of both projects, given the centrality of such notions to OS and to the various permutations of OEF.

I said in Chapter 4 that I would try to clarify the pretheoretical concept of faculties by holding it up against the cognitive scientific concept of cognitive systems. As I mentioned there, I think that this is perhaps the most fruitful approach. This means, however, that I will have to develop the cognitive scientific notion of systems in detail and pretty much from scratch. There is a good deal of discussion in cognitive science about what cognitive systems there are, but very little about what cognitive systems are.\(^1\) A large part of the present chapter is devoted to explicating this scientific concept, and I should justify myself for spending so much time on it.

First of all, I assume that our pretheoretical intuitions about whether or not

\(^1\) I do not mean to imply that the issue has been completely neglected; Bedford (1997) and Kosslyn & Koenig (1992), for example, have addressed this general question, but I am looking for something more precise than they attempt. Fodor (1983) and others have had a lot to say about modules, which comprise one particular kind of cognitive system, but Fodor takes the more general concept of cognitive systems as primitive, and besides, I am looking for more precision than he offers.
something constitutes a faculty are often not very clear. This suggests that the underlying conception of faculties is itself somewhat vague. Since one of the desiderata of the descriptive epistemological project is to retain (even predict) the vaguenesses and indeterminacies inherent in our intuitive judgments, a descriptive epistemological account of faculties should be vague as well. This means that there is not all that much work to do regarding this project. Normative epistemology, on the other hand, requires something far more precise. One of the goals here is to eliminate whatever vagueness occurs in the descriptive account. For this reason alone, a normative epistemological account will have to be more detailed than a descriptive epistemological account.

Secondly, the success of the normative epistemological account I have offered depends on what opaque systems there happen to be. There are a number of researchers who doubt that, as a matter of empirical fact, there are many, if any, modules in Fodor's sense (e.g., Elman, et al., 1996; Karmiloff-Smith, 1992), and I have some sympathy with this position. Therefore, it is especially important to distinguish cognitive systems from modules, so that the contentiousness of the modularity doctrine does not infect the normative epistemological account already offered. By developing the notion of cognitive systems in detail and in response to empirical findings and practices, as well as by explicitly contrasting it with a conception of modularity, I hope to show how there might be many more opaque systems than there are modules.

Finally, the notion of cognitive systems is of independent interest to cognitive science. Debates about what systems there are could benefit greatly from the development of a precise understanding of what a cognitive system is. And it is largely for this reason that I will attempt to develop the notion on its own strength, in the sense that all the evidence I will use in support of my theory of what a cognitive system is will be cognitive
scientific evidence, not epistemological evidence. In the end, however, I think that this also helps to show that the notions of cognitive systems and faculties -- which the epistemological accounts rely upon -- are robust concepts in their own rights and have not simply been tailor-made to get the cases right.

In Section 1, I develop a theory of what a cognitive system is. I start with an intuitive characterization of the target before turning to a specific methodological paradigm used in empirical studies of systems, using the assumptions implicit in that methodology to develop a more precise definition of cognitive systems. In Section 2, I compare this notion of cognitive systems to something like Fodor's notion of modularity to show how the former differs from, clarifies, and improves on the latter. In Section 3, I compare cognitive systems to cognitive faculties and explain the sense in which the former is just an explication of the latter. This distinction between cognitive systems and cognitive faculties makes it possible to answer an objection that was left over from Chapter 4. I do this in Section 4. Finally, in Section 5, I offer some brief suggestions about how the resulting conceptions of faculties and cognitive systems might serve as the basis for pursing an elaborative epistemological account of the P/N distinction.

1. A Conception of Cognitive Systems

To get started, consider a somewhat influential distinction and analogy due to Aleksandr Luria:

Those investigators who have examined the problem of the cortical "localization" of elementary functions by stimulating or excluding local brain areas have understood the term "function" to mean the function of a particular tissue. . . . When we speak of the "function of digestion" or "function of respiration" [on the other hand] it is
abundantly clear that this cannot be understood as a function of a particular tissue.

(1973, p. 27)

'Function', as it appears in 'a function of the liver' is quite different from 'function' as it appears in 'the function of respiration'. Luria calls the former a "simple function" and the latter a "functional system" (the terminology will prove unfortunate, since what he is calling a "functional system" is directly analogous to what I want to deny is a cognitive system; for now I will maintain his terminology). A diagnostic feature of a "functional system", according to Luria, is that its behavior is subserved by a variety of (possibly disjoint) mechanisms:

For instance, if the principle group of muscles working during respiration (the diaphragm) ceases to act, the intercostal muscles are brought into play, but if for some reason or other they are impaired, the muscles of the larynx are mobilized and the animal or person begins to swallow air, which thus reaches the alveoli of the lung by a completely different route. The presence of a constant (invariant) task, performed by variable (variative) mechanisms, bringing the process to a constant (invariant) result, is one of the basic features distinguishing the work of every "functional system". (1973, p. 28; italics in original)

But what does this have to do with cognitive systems? If we can appropriate Chomsky's (1980) phrase without importing his doctrines, a cognitive system is basically a "mental organ"; the task of carving the mind into systems is, roughly, the task of individuating mental organs. And it is here that Luria's illustration is useful. One important reason why we should not claim that there is a respiratory organ is that respiration is something that can be done by a number of distinct (though partially overlapping) complexes of tissue. The concept of a respiratory organ would have to be a
disjunctive concept: the lungs and the diaphragm or the lungs and the intercostal muscles or the lungs and the larynx, etc.. This has the feel of being gerrymandered and unprojectible (in the sense of Goodman [1955]) and thus should be avoided if possible. The notion of a circulatory organ, on the other hand, does not seem as bad, for even though it is a conjunctive notion (it includes heart, veins, arteries and capillaries), there is no alternative way to circulate blood (at least, for example, in mammals). Since it is the same set of mechanisms being used every time, any properties we observe it to have in one creature at some time should reasonably be expected to appear in similar creatures at other times.

1.1 Intuitive Distinctions

With this very general and vague notion of simple functions in mind we can turn to some particularly cognitive examples, for it is roughly this notion of cognitive systems as mental organs and (mental) organs as simple functions that at least some of cognitive science (and cognitive neuroscience in particular) has in mind when talking about cognitive systems.\(^2\) Again, Luria’s terminology is misleading here, for what he calls a “functional system” is something that many cognitive scientists wouldn’t even consider a system at all. There is a sense in which what cognitive science has taught us about memory, for example, indicates that there is no such thing as a general memory system. There are, perhaps, an episodic memory system, an iconic memory system, a “medium term” (hippocampal) memory system and so forth, but these systems do not sum together to comprise a single overarching system that handles memory. Remembering is a task that is, in Luria’s terms, “performed by variative mechanisms”; there is a disjoint set of mechanisms that come into

\(^2\) Kosslyn and Koenig (1992), for example, are explicit about this; they cite Luria approvingly.
play in different instances, depending on the precise nature of the mnemonic task in question. And it is because of this disjointness that there is not one overarching memory system.

Sometimes, however, several smaller systems do sum together to yield one larger system. Compare, for example, the account of memory just sketched with a particular account of one kind of visual object recognition. According to an influential view (Biederman, 1990), there are several subsystems involved in intensity-based visual object recognition: low level systems that extract edge information, higher level systems that determine arrangements of lines and vertices from this information, still higher level systems that compose viewpoint-invariant representations from these. But all of this is consistent with this being a visual object recognition system, because each of these subsystems is engaged in a different task, and because the task of visual object recognition is analyzable into these subtasks.3 This is not a case of varying mechanisms (mechanism A on some occasions, mechanism B on other occasions) subserving a single task. The task of remembering might be divided into the tasks of episodic remembering, semantic remembering, and so on, but this division is a matter of resolving an ambiguity in the term ‘remembering’ and not a matter of task analysis.

Obviously, the details of Biederman’s story are not important here. It does not matter for the present purposes whether, for example, this sort of visual object recognition actually involves the construction of viewpoint invariant or viewpoint dependent (e.g., Tarr & Büllthoff, 1995) representations. The point is that, whatever theory is right, there is presumably some task (perhaps it would be better labeled “visual, intensity-based, basic-

---
3 I do not mean, and do not mean to suggest, that there is only one visual object recognition system. The point here is merely to illustrate the difference between subsystems of a single system and disjoint mechanisms.
level object categorization for non-face stimuli”, or something similar) that analyzes into several subtasks, where these subtasks are not alternative ways of getting the same result, but are necessary stages of the overarching task. It is in this way that visual object recognition (for a certain level of categorization and for a certain class of stimuli) comprises a system in a way that memory (in general, without qualification) does not. The variety of mechanisms that subserve the task of remembering do not constitute subsystems of a larger memory system, at least not in the same sense that the mechanisms involved in visual object recognition do. I will restrict my use of the terms ‘system’ and ‘cognitive system’ to the sense gestured at above, that is, to the mechanisms that subserve what Luria calls “simple functions”.

One useful way of looking at this is in terms of “boxologies”. One goal of cognitive science is to carve the mind into a number of systems, where the systems are represented as boxes. Visual object recognition, according to the above sketch, could be drawn roughly like this:

![Diagram of visual object recognition]

This is not, of course, the only way in which boxes are used. Philosophers, for example, following Schiffer (1987), often draw in a “belief box” in diagrams of very gross level cognitive structure, although no one (I hope) thinks that there is a single system devoted to belief-formation. This sort of functional individuation of boxes is much more coarse-grained than the one I have in mind here. Another use of boxes, which is orthogonal to the one presented here, is often used by neuroscientists to represent the cytoarchitectonic (though perhaps not the functional) features of the brain. The labels on the boxes indicate how these boxes are being used. If the boxes contain physical level descriptions like ‘V1’ and ‘MT’, it is a neuroanatomical boxology; if they contain task descriptions (i.e., semantic level descriptions), then it is a functional boxology.
The arrows indicate the flow of information from lower to higher levels, as the theory adumbrated above is a model of bottom-up processing. The box labeled 'Visual Object Recognition' may feed into other boxes and/or be included in a larger box, the smaller boxes here might contain even smaller boxes as parts, and so forth. There is no reason why we could not draw a box around the first two boxes if we wanted to talk about the system that extracts vertex information from retinal stimulation rather than from edge representations. (I am assuming that the edge detection system takes its input from the retina.) This points to an ambiguity in the labeling of boxes, for it would be natural to call this system a vertex extraction system, although that label has already been applied to the second box included in the larger box:

![Diagram](image)

There is nothing wrong with this as long as we distinguish broader from narrower conceptions of systems. Conceived narrowly, the vertex extraction system is the small box on the right inside the larger box; conceived broadly, the vertex extraction system is the larger system, which includes an additional system necessary for the task of vertex extraction. Similarly, in diagram (1) above, the last box, labeled "matching", could have been labeled "visual object recognition", narrowly conceived, of course. The distinction between broad and narrow conceptions of systems parallels and is parasitic on a distinction between conceptions of tasks: the task of vertex extraction, broadly conceived, includes the subtask of edge detection, while the task of vertex extraction, narrowly conceived, does not. I will make all this more precise in section 1.2; a rough and intuitive understanding
should suffice for now.

Now, not just any combination of subsystems constitutes a system. It would be unnatural, for example, to enclose the first two and the last two sub-boxes in (1), above, in a single box, omitting the line completion system, as follows.

\[
\begin{array}{c}
\text{edge detection} \rightarrow \text{vertex extraction} \\
\text{?} \\
\text{geon recovery} \rightarrow \text{matching}
\end{array}
\]

It is unclear what the resulting "system" would be -- hence the dotted line and the question mark. One of the problems is that there is no arrow from vertex extraction to geon recovery. If the earlier diagram is correct, and geon recovery requires line completion, then there is no system that computes geons from mere vertex representations. The above does not look like a system at all, but rather, a gerrymandered collection of parts of a system. If this would count as a system, then it would be hard to see what would not count as a system. But clearly not everything counts as a system, at least not in any interesting sense of the term.

Similarly, suppose that working memory is divided into (at least) an articulatory system (Baddeley & Hitch, 1977) and a visuo-spatial system (Baddeley, 1986), where the former handles the short-term retention of verbal stimuli and the latter, visual stimuli. Suppose further that these two do not interact in any way, either with each other or with any other kind of working memory. Would the following be a representation of a working memory system?
Again, the answer seems to be no. If we allowed that this were a system, it is hard to see what would not be a system; any set of systems would itself be a system. The idea of a short-term memory system seems more natural to us than many of the "systems" we could put together in this way, because short-term memory seems intuitively like a coherent category. If the search for systems is to be an empirical issue, however, we cannot take such intuitions too seriously. Denying that there is a short-term memory system does not mean that there is no such thing as short-term memory; of course there is. What it means is that the systems that are involved in short term memory bear (or, rather, fail to bear) a certain causal relation to each other.

This notion of a causal relation needs to be stressed, for it places important restrictions on our use of the term 'function'. Sometimes this term refers to teleology, sometimes to a mathematical notion of a mapping, and sometimes to behavior in general. Obviously, not all of a mechanism's behavior is relevant from the standpoint of determining whether it realizes a system or not. Cognitive systems are supposed to be functionally individuated, but the relevant notion of function must be narrower than that of behavior. It is only the cognitive behavior of the underlying mechanism that matters. Mechanisms, qua chunks of brain, will certainly engage in behavior that is not cognitive: they will displace water, give off a certain odor, and so forth. If this sort of behavior is characterized as a mapping (e.g., taking being shot out of a cannon as an argument and returning describing a parabola as a value), then it is clear that brains and their parts will
produce a number of mappings that are irrelevant from the perspective of cognitive science. The relevant behavior they exhibit should be characterized by the cognitive mappings they produce, i.e., by those functions (in the mathematical sense) they compute, which take cognitive states as their arguments, values, or both.5

This alone, however, does not give us any grounds on which to individuate systems. If all we knew were the (cognitive) mappings, then the best we could do would be to lump input/output pairs together based on our commonsense views of what sorts of mappings constitute coherent categories. But this fails to distinguish memory from visual object recognition in the right way. Carving the mind into systems, on this approach, would not be a particularly empirical endeavor -- it certainly would not be the sort of endeavor that required expensive imaging equipment or the lesioning of animals. From the standpoint of folk psychology, in fact, the notion of a memory system is far more coherent than that of a visual, intensity-based, basic-level object categorization system for non-face stimuli. But there is a sense of "cognitive system", according to which the latter appears more likely to be a genuine system than the former.

1.2 Double Dissociations and a More Precise Conception of Systems

The visual object recognition system under discussion here demonstrates a certain causal interdependence among its parts, while the non-system of memory does not. To borrow an illustration that Haugeland (1978) uses for a somewhat different purpose, the parts of a car's engine are causally interdependent in the sense that neither the carburetor nor the distributor nor the spark plug has any independent contribution to make to the

---

5 I do not intend to defend any particular view as to which states count as cognitive states. The view I take to be standard is espoused by Fodor (1975) and Pylyshyn (1984), but I would be happy to entertain an alternative account.
turning of the crankshaft. It is only the various parts working concertedly that cause the crankshaft to turn.

In the last section, I compared some likely candidates for systems and for non-systems in order to illustrate this causal interdependence. This notion of causal interdependence (or rather the absence of such interdependence) provides the foundations for the neuropsychological methodology of double dissociations, and in this section I want to explicate this notion of causal interdependence in light of this methodology. Note that I have no intention of operationalizing cognitive systems in terms of double dissociations. I want claims about the existence of certain cognitive systems to be made reasonable by findings of dissociations, but not analytically implied by them. I focus on double dissociations mainly because this is the oldest and most familiar approach to distinguishing cognitive systems. (Another reason will become clear in Section 3).

The way double dissociations are typically used is as follows: if damage to a particular area causes an impairment with respect to task A but leaves performance on task B intact, this is a single dissociation. Such a dissociation suggests a distinction between systems for tasks A and B, for it implies a certain amount of non-interdependence. According to methodological orthodoxy, however, a far stronger case can be made upon discovering a double dissociation; single dissociations do not indicate that different systems subserve the two tasks, for it might just be that task A is simply more demanding in the sense that performance on it degrades less gracefully than performance on task B, even though it is the same system that performs both tasks. If another dissociation is found, however, with performance on A spared but performance on B impaired, this constitutes a double dissociation and eliminates the possibility that A is simply more demanding. Such evidence is typically taken to imply that A and B are subserved by distinct systems.
To make this more concrete, let us turn to a well-known study of primate vision. Ungerleider, Mishkin, and others (Ungerleider and Mishkin, 1982; Mishkin, Ungerleider and Macko, 1983) trained monkeys on two tasks. In the landmark discrimination task, food was hidden in a well, and a cylinder was placed near the well that contained the food. The monkeys were to use the landmark (the cylinder) to locate the food. In the delayed non-match to sample task, monkeys were shown pairs of objects, one of which had been seen before and one of which had not; their task was to choose the new object. Monkeys with bilateral lesions to posterior parietal cortex were impaired on the landmark discrimination task but not on the delayed non-match to sample task. Monkeys with bilateral lesions to inferior temporal cortex, on the other hand, showed the opposite symptoms; they were impaired on the delayed non-match to sample task but not on the landmark discrimination task. These researchers concluded that there are at least two distinct visual systems, one in the posterior parietal lobe that computes "where" information, and another in the inferior temporal lobe that computes "what" information.

There is, of course, some controversy concerning whether Ungerleider and Mishkin properly identified the systems as "what" and "where" systems. Goodale & Milner (1992), for example, argue that these systems are really visual perception and visually-guided reaching systems, respectively. What does not seem to be in dispute, however, is that the double dissociations indicate the presence of two distinct systems. Similar uses of the double dissociation paradigm abound in the literature on the neuropsychology of memory (e.g., Squire & Butters, 1992). In fact, it is largely because of these dissociations that we know that the different memory systems are not subsystems of a single memory system.

The sort of causal interdependence that is involved in systems (the lack of which is
indicated by dissociation) still needs to be spelled out in more detail. Let me introduce a bit of terminology here that will become important later. Up until now, I have been playing fast and loose with the distinction between cognitive and implementational levels. From here on out, I will use the term 'system' to refer to virtual machines that are realized (and possibly multiply realizable) in physical objects. I will use the term 'substrate' to refer to anything at the physical level of description (a.k.a., the realization, or implementational level), an object or part of an object which may or may not realize a system. Thus understood, an important goal of cognitive neuroscience is to determine which substrates realize which systems. I will retain the term 'mechanism' to refer indiscriminately to something at either level, when it is not important to distinguish the cognitive from the implementational level.

At least in cases where a system is comprised of subsystems, the above examples indicate that these subsystems must causally interact in some yet to be specified way. (This interaction will, of course, be a consequence of the interaction of their underlying substrates.) Double dissociations are used to infer that some putative system (e.g., memory) is too "big" to be a genuine system. But there are other putative systems which may be too small to be genuine systems; the task they perform may be just part of the task of a primitive system, i.e., a system that is not composed of other systems. Some decent prima facie candidates for primitive systems might include any of the small boxes drawn above: vertex extraction, geon recovery (construed narrowly), visuo-spatial working memory, and the rest. Unlikely candidates include, e.g., speech processing and the

---

6 The existence of double dissociations tell us little or no more than this. While such findings may indicate that two tasks are subserved by different systems, the existence of double dissociations tells us nothing about what tasks are subserved by the same system. For this, we need some reason for thinking that the two tasks will not dissociate. Obviously, such negative evidence is harder to come by.
recognition of my grandmother’s face. These last two seem unlikely for very different reasons: the former is probably too big to be primitive, while the latter is probably too small to be a system.

It is easy enough to come up with many more examples of tasks that appear to be too small, or narrow, to delineate a system: a system for reading the word ‘Frenchie’, a my-left-arm-is straight-in-front-of-me proprioceptive system, etc.. This narrowness, however, is not a necessary or intrinsic feature of the task in question; while it seems unlikely that we humans have owl detection systems, it would not be very surprising at all if field mice did. Whether a task is too narrow to delineate a system depends on the cognitive organization of the organism; it cannot be read a priori off of the nature of the task itself. So what is it that these tasks share that is not shared by, e.g., line completion? What makes these tasks appear too narrow is that they do not seem to be isolable. The reason that we do not take the possibility of a grandmother detection system seriously is that we are convinced that the very same substrates that underlie grandmother recognition also subserve other tasks. As a matter of empirical fact (or so it certainly seems) there is no such thing as grandma-agnosia.

To return to an earlier example, suppose that the ventral stream from primary visual cortex to the inferior temporal lobe computes not only the identification of objects, but also certain information about the location and size of the objects (Goodale, 1995; Goodale & Milner, 1992). Would it still be appropriate to call the ventral stream a “what” system? According to Ungerleider and Mishkin’s famous proposal, “what” information and “where” information are computed separately in such a way that the brain later has to solve a binding problem (Crick & Koch, 1990): the position and the identity of the object must come together to yield a representation of an object of-kind-\(K\) at-location-\(L\). But extending
the above supposition, suppose that the ventral system's computation of what and where
information occurs together, in such a way that there never is a binding problem that has to
be solved. Now it would definitely seem wrong to call this a "what" system, and in the
same way that it seems wrong to posit a grandmother recognition system. There would
not be a "what" system and a "where" system; instead there would be a "what and where"
system. This would be a case where the mechanisms (the substrates) that subserve the
computation of object identity also subserve other tasks.

This general idea, however, needs to be narrowed down somewhat, for the brain
computes information concerning the identity of faces, yet the brain obviously computes
other things. That is, there is a substrate (the brain), which subserves face recognition, but
which subserves other tasks as well. Yet we would not want to say merely on these
grounds that there is not a face recognition system. What has gone wrong here is that the
relevant substrate was not specified narrowly enough. The fact that there is a substrate
(viz., the brain) that performs face recognition but also does other things does not mean
that there is no face recognition system, for it does not imply that there is not another,
smaller substrate which performs face recognition and only that. To say that there is a face
recognition system is to say, roughly, that there is a substrate which computes information
about the identity of faces, and that this substrate is both independent, in the sense that its
performance does not require any other substrates, and specialized, in the sense that it does
not perform any other tasks.

The two features of independence and specialization have very different roles to
play. Every independent substrate realizes a system, and every system specializes in
something or other. In a worst-case scenario, there is only one independent substrate (the
brain) and it specializes in cognition in general. This may sound like a degenerate sense of
'specialize', but I do not mean for the notion to be very robust. The notion of independence is far more important than that of specialization. The latter only determines what we call the system; it is the former that actually serves as the basis for individuating systems. Independence yields the boxes themselves; specialization only provides labels for them.

This needs to be made more precise, and we should start with a clearer understanding of tasks. The individuation of tasks can be either fine-grained or course-grained. To take a prosaic and non-cognitive example, suppose Ed and Karen both bake a cake. Their results are qualitatively identical, but Ed baked his from scratch, while Karen baked hers from a prepared mix. Here we would want to say that there is a sense in which they both performed the same task (viz., baking a cake), but another sense in which they performed different tasks (since Ed did things that Karen did not do). This is because tasks are typically, but not always, named for their end results. A mechanism that computes parse trees from written text performs the same task, in a sense, as one that computes parse trees from acoustic speech signals; but in another sense, they perform different tasks. A course-grained individuation of tasks only considers the outputs, while a fine-grained individuation of tasks considers the starting points as well. I will rely on a fine-grained individuation of tasks in what follows unless I explicitly state otherwise.

Since the only tasks I will be concerned with are cognitive ones, let me define a task, for the present purposes, as a cognitive function, in roughly the sense suggested in section 1.1. However, this notion of cognitive functions itself needs a bit of fine-tuning. I will define a strictly cognitive function as a mathematical input/output mapping that takes cognitive states as its arguments and its values. A transduction function I will define as a mathematical input/output mapping that takes (a) cognitive states as its arguments and
motor responses as its values, or (b) sensory stimulation as its arguments and cognitive states as its values. $F$ is a cognitive function if and only if $F$ is either a strictly cognitive function or a transduction function. To perform a task is simply to compute the cognitive function that is that task.

Since tasks are just functions, a task can be represented as a set of ordered input/output pairs. Now there are two importantly different ways to divide up tasks; these were alluded to in 1.1: a task can be partitioned into parts or analyzed into subtasks. A part of a task, as I will use the term, is simply a subset of the set of input/output pairs which constitute that task. A subtask of task $T$ (relative to some set of mechanisms that compute $T$), on the other hand, is a function which is not identical to $T$ but which gets computed (by those mechanisms) in the process of computing $T$. Returning to memory and visual object recognition, recovering semantic memories is a part of the task of remembering, although presumably not a subtask, and vertex extraction is a subtask of the task of forming basic level category representations from retinal stimulation, although it is not a part of this task. What parts of a task there are can be determined solely from a specification of the task itself, i.e., from the set of input/output mappings. What subtasks there are can only be determined by reference to the algorithm used. Consider, for example, the task of multiplication. Some particular multiplier might be using a partial products algorithm, in which case, it performs the subtask of computing the product of the last digits of both numbers. If it is using a successive addition algorithm, on the other hand, it will not perform this subtask and instead the task it performs will be analyzable into a quite different set of subtasks. Thus, when speaking of subtasks, we need to (at least tacitly) relativize the task in question to some particular class of mechanisms for computing the task.

The restriction to input/output functions in the definition of tasks is intentional, for
without it, it is impossible to distinguish parts from subtasks in the above manner. Suppose we have a system that takes representation $R1$ as input and tokens $R2$ as an intermediate stage on the way to producing $R3$ as an output. Suppose that although it tokens $R2$, it does not output $R2$. Such a system would compute the function that maps $R1$ to the set $\{R2, R3\}$, but since it does not output $R2$, the input/output function that it computes maps $R1$ only to $R3$, and not to either $R2$ or to $\{R2, R3\}$.

This notion of analysis into subtasks provides a more precise understanding of the distinction between broader and narrower conceptions of tasks. If $T$ and $T^*$ are tasks that have the same outputs, then $T$ is narrower than $T^*$ (relative to some mechanism) if and only if $T$ is a subtask of $T^*$ (relative to that mechanism), i.e., the computation of $T^*$ (by that mechanism) includes the computation of $T$ as a stage of computation. $T$ is broader than $T^*$ if and only if $T^*$ is narrower than $T$. This leaves the notions of broader and narrower undefined for tasks that do not have the same outputs (and even some that do), but this is as I think it should be. If $T$ and $T^*$ do not have the same outputs, then neither is broader or narrower than the other. Broad and narrow conceptions of systems follow directly from this. Systems derive their labels from the outputs of the tasks they perform, and this is why there is an ambiguity between the broader and narrower conceptions of systems. Systems $s$ and $s^*$ (for the sake of clarity, here and henceforth I use capital letters to denote substrates, and lower case letters for systems or candidate systems) are systems for broader and narrower conceptions of the same task (individuated at a coarse grain), respectively, if

---

7 This is not exactly the same as what we started with, which was two conceptions of the same task. What we have now is a relation between two different, albeit output equivalent, tasks. What I am proposing is that we replace the intuitive notion of different conceptions of the same task (which relies on a coarse-grained individuation of tasks) with a specification of a relation between two tasks, where these are subject to fine-grained individuation.
and only if the task that \( s \) performs is broader than the task that \( s^* \) performs (with tasks here individuated at the fine grain, of course).

The notion of partitioning tasks into parts plays a different and more important role, for it occurs in the general definition of a system. I will say that a substrate is \textit{isolable with respect to task} \( T \) if and only if it performs task \( T \) and could do so even if nothing else computed any (cognitive) functions. A substrate \textit{specializes in task} \( T \) if and only if \( T \) is an exhaustive specification of the cognitive input/output function that that substrate computes. I will call a task \textit{unitary with respect to substrate} \( S \) if and only if \( S \) performs that task and no proper part of \( S \) is isolable with respect to any part of that task. Intuitively, if \( T \) is unitary with respect to \( S \), then nothing less than the whole \( S \) can perform anything less than the whole \( T \). I propose that:

- A substrate \( S \) realizes a system for task \( T \) if and only if (i) \( S \) is isolable with respect to \( T \), (ii) \( S \) specializes in \( T \), and (iii) \( T \) is unitary with respect to \( S \).

Primitive systems are easily defined in terms of the more general notion of systems: if \( S \) realizes a system for \( T \) and no proper part of \( S \) realizes a system for any subtask of \( T \), then, and only then, \( S \) realizes a primitive system for \( T \).

The isolability constraint rules out things like, e.g., the output layer of a connectionist network. The substrate that realizes just the output layer does produce outputs, in the presence of other substrates (viz., those that realize the input and/or hidden layers), but it cannot compute a (cognitive) function in the absence of such substrates. The specialization constraint prohibits things like grandmother recognition systems (provided that we are right about how grandmother detection works): any substrate that computes a function whose outputs are grandmother representations also computes other functions with different outputs. The unitariness constraint precludes things like a memory system
(or a cognition system, whose substrate is the whole brain), for such tasks can be partitioned into parts, some of which are computed by isolable substrates.

Now this is not yet a definition of cognitive systems; rather, it is a definition of something's realizing a cognitive system. I think that something's realizing a cognitive system is more fundamental than something's being a cognitive system, and as a result, a definition of cognitive systems follows fairly straightforwardly from the above definition:

• $s$ is a cognitive system for task $T$ if and only if there is some substrate $S$ that realizes a cognitive system for task $T$ in virtue of realizing $s$.

In most of what follows, I will concentrate for the most part on the more fundamental notion of realizing a system.

Earlier, I defined a part of a task as a subset of the set of input/output pairings that constitute that task, rather than as a strict subset. There is a reason for this. Consider a pair of redundant systems that perform the same task, but which function independently of each other:

\[
\begin{align*}
\text{a : task } T & \quad \text{(5)} \\
\text{b : task } T & \quad c
\end{align*}
\]

In this case, $a$ and $b$ together do not constitute a system. The present account captures this in the following way: $C$ (the substrate of $c$) specializes in and is isolable with respect to $T$, but $T$ is not unitary with respect to $C$. This is because $B$ (the substrate that realizes $b$) is isolable with respect to $T$ and $T$ is a part of $T$, since 'part' is defined in terms of subsets rather than strict subsets. And since $B$ is a proper part of $C$, $T$ is not unitary with respect to $C$. The same holds for $A$.

The distinction between the two different kinds of cognitive functions provides for
a distinction between three interestingly different kinds of cognitive systems. We can distinguish between what I will call "strictly cognitive systems", "transduction systems", and "transducers". \( s \) is a strictly cognitive system if and only if \( s \) computes a strictly cognitive function; \( s \) is a transduction system if and only if \( s \) computes a transduction function; \( s \) is a transducer if and only if \( s \) is a primitive transduction system. So, for example, any primitive system that takes retinal stimulation as its input would constitute a transducer. Edge detectors, presumably, are transducers in this sense, provided that the edge detectors are primitive systems. A system that takes retinal stimulation and returns, say, geon representations, or object identifications, is a transduction system, but presumably not a transducer (since such systems are presumably not primitive). And a system that computes object identification from edge representations would be a strictly cognitive system.

Some consequences of the above definition are worth pointing out. It is possible to have a number of nested systems, like the following:

\[ \begin{array}{c}
\text{a} \\
\text{b} \\
\text{c} \\
\text{d}
\end{array} \]

where \( d, e, f, \) and \( g \) all produce the same output. If \( d \) performs, say, face recognition from some relatively high-level representations, then \( d, e, f, \) and \( g \) would all be face recognition systems, conceived more or less broadly; we get a hierarchy of progressively broader and narrower conceptions of a system. In this sort of case, \( d, e, f, \) and \( g \) will all have the same name, since they all produce the same output. If we need to distinguish them, we will have to specify the task they perform by a fine-grained individuation of tasks, i.e., we would
have to specify the cognitive input/output functions they compute.

Not only is the nesting of systems possible, it seems likely that virtually any time one system feeds directly into a second system, then these two systems constitute a third system. In addition, it seems that if \( x \) is a putative system which contains nothing except for three systems, \( I, m, \) and \( n \), and the operation of \( n \) depends on the combined contributions of \( I \) and \( m \) (\( n \) cannot perform its task without receiving inputs from both of the other two systems), then \( x \) is a system. (See illustration (7), below.) However, we would not want to say that the set of \( I, m, \) and \( n \) comprise a system merely because \( n \) receives inputs from \( I \) and \( m \); it is essential that \( I \) and \( m \) operate in conjunction. If \( I \) and \( m \) work independently with respect to \( n \), in such a way that the operation of either is sufficient for \( n \) to perform part of its task, then the substrate that realizes \( n \) and that system will be isolable with respect to some part of the task that the combination of the three performs; thus, this task is not unitary with respect to the combination of the three, and hence the combination of the three does not realize a system.

Our scheme of boxes and arrows can reflect this if we use converging arrows to represent joint contribution and parallel arrows for independent operation:

Thus, while \( x \) is a system, \( y \) is not. Instances of systems of the form of \( x \) might include a system that takes visual information from what and where systems and binds them into a unified percept. If there is a distinct system responsible for the McGurk effect (McGurk & McDonald, 1976), it almost certainly has this structure. Radeau (1994) goes so far as to claim that there is a Fodorian module (and \textit{a fortiori} a system, we might add) dedicated to
the pairing and integration of visual and auditory stimuli. In this sort of case, if $n$ is the binding system (narrowly conceived) and $l$ and $m$ are the what and where systems, respectively, we could view $x$ as the what-where binding system, more broadly conceived. Examples of things like $y$ are probably also quite common: higher level reasoning systems, for example, will take inputs from different sources, and these inputs are likely to be independent of each other in the sense that the reasoner can do its work as long as one system or another feeds it premises; it does not require more than one source of premises. Schacter's (1989) conscious awareness system, analogous to box $n$, takes input independently from the systems that feed into it, and thus provides another instance of $y$.

The present account clearly implies that there will be a great many non-primitive systems if there are more than a handful of primitive systems. (Non-primitive) systems can overlap in a number of ways: they can be nested (as in (6) above), they can converge (as in (7) $x$ above), they can diverge, like so:

They can even, I suppose, cross paths, at least in principle:

Many of the resulting non-primitive systems will not be very interesting, and counting the number of nested systems involved in some task will not be very informative. Still, the
present proposal provides some principled grounds for ascribing system status to some, but not all, candidates, and it offers some hope for developing what we might call a "syntax of boxes", a set of principles describing which combinations of smaller systems yield larger systems.

2. Systems and Modules

I think that it is easier to see what is involved in the current understanding of cognitive systems if we compare this notion to what I suspect is a more familiar one: the Fodorian conception of modularity. According to Fodor (1983), the distinguishing characteristics of modules are speed, mandatory operation, shallowness of outputs, inaccessibility of interlevel representations, neural localization, characteristic ontogeny, characteristic pathology, informational encapsulation, and domain specificity.

Now I am going to need a fairly precise account of what a module is in order to compare it to the present understanding of systems. To this end, I will take the nine characteristics just mentioned and consider them to be individually necessary and jointly sufficient for a specific sort of module: an "F-module". Although I think that this is a fairly common understanding of modularity, it is actually not Fodor's official view: he claims that "[t]he modularity of the input systems consists in their possession of most or all of the properties" just mentioned (1983, p. 47; italics added). He also denies that his characterization of modules is supposed to constitute a definition of modules (p. 37). Since the account of cognitive systems offered here is intended to serve as a definition (though not necessarily as a piece of conceptual analysis), I want to compare it to something that is a definition. It is for this reason that I am making the possession of the nine aforementioned traits definitional of F-modularity. This, I think, is close enough to
Fodor's own view to vindicate the use of the 'F' in 'F-module'.

Some of the aforementioned features have received more attention than others. In particular, commentators have concentrated on the characteristic ontogeny (modules are innately specified), the inaccessibility of interlevel representations (the computations a module performs on the way toward producing an output are not available to introspection), informational encapsulation (modules do not have access to all of the agent's beliefs and desires), and domain specificity (modules are special-purpose mechanisms). Since these are the most important features of modules, I will concentrate on these notions in what follows. The main purpose here is to clarify the present notion of cognitive systems, since it plays such a central role in the account of the P/N distinction, but since the notion of cognitive systems is of independent interest, I will also point out a few of the advantages (conceptual and empirical) that the present notion has over that of F-modularity (and, to a somewhat lesser extent, Fodor's non-definitional understanding of modules).

2.1 A Special Case of Cognitive Systems

With the exception of neural localization, none of the above characteristics of F-modules are required by the more general notion of cognitive systems, although, of course, the definition of systems is consistent with a system having any number of these properties. The last two features listed above, informational encapsulation and domain specificity, have close analogues in the present criteria of isolability and specialization,

---

8 The literature on modularity is far too large to be referenced even nearly exhaustively. For some representative material, a number of relevant papers are conveniently collected in Garfield (1987). For some more recent efforts, see Samuels, et al. (forthcoming); Segal (1996); Sperber (1994).
respectively, but isolability and specialization are weaker in important ways than the notions of informational encapsulation and domain specificity.

I will begin with informational encapsulation. Although we can understand informational encapsulation as a monadic property (some systems are encapsulated, *simpliciter*), there is also an important *relational* understanding of encapsulation, according to which some system is encapsulated *from* some other system. We can say that \( s \) is encapsulated from \( r \) if and only if \( s \) has no access to any of the representations that \( r \) computes; informational encapsulation *simpliciter* can presumably be cashed out in terms of relational encapsulation, though not vice versa.\(^9\)

Consider, for example, a long-standing debate between theories of language comprehension. On a strictly bottom-up theory (e.g., Forster, 1976, 1990), we may have a syntax F-module with one-way connections into a semantics F-module; i.e., the sentence is parsed first, perhaps in several different ways, and candidate parse trees are sent to the semantics processor, which attempts to assign meaning to the sentence, eliminating some of the candidates it receives and sending the remaining possibilities on toward central processors. On this view, the syntax system is encapsulated from the semantics system: the former does not receive any input from the latter. One kind of competing theory, on the other hand, insists that there is a great deal of top-down processing from semantics to syntax (Cole & Jakimik, 1980; Seidenberg & McClelland, 1989). I think that we would all agree that this would entail that there is not an F-*module* for syntax (because syntax, on this

---

\(^9\) A good start might be to claim that a system is encapsulated *simpliciter* if and only if it is encapsulated from all higher-level systems (though this would require some independent understanding of which systems are higher-level than which other systems). Since my interest in F-modules here is only by means of contrast to systems, I will not pursue these details. I should point out that, although Fodor's discussion of informational encapsulation is largely concerned with the monadic property, it is clear that he does recognize encapsulation as being also a relational property (e.g., 1983, p. 69).
view, is not informationally encapsulated from semantics), but would it mean that there is no syntax system?

Whether there is a syntax system in this case depends, I think, on the nature of the top-down influence. To start with a fairly easy case, suppose that the syntax and semantics mechanisms interact completely -- syntactic parsing relies essentially on semantic analysis and vice versa. I think that in this case, it would be pretty clear that there is no syntax system, just as there is no system for recognizing the word ‘Frenchie’. Certainly, we could draw boxes labeled ‘syntax’ and ‘semantics’ with two-headed arrows between them, but what would be the rationale for drawing two boxes in the first place? That is, why not just have one box labeled ‘syntax and semantics’? In this case, where the two mechanisms are so intertwined, it seems that the only motive for drawing two boxes is the fact that parsing and assigning meaning strike us as distinct tasks. Of course these are conceptually distinct, but the question is whether there is any psychological distinction between them. Here the answer seems to be no. Any reason for thinking there is this sort of interaction between the two boxes is a better reason for thinking that it is just one box after all.

In Fodor’s words, “it is a point of definition that distinct functional components cannot interface everywhere on pain of their ceasing to be distinct” (1983, p. 87; italics in original). But how much interaction between distinct systems is tolerable? Informational encapsulation, strictly speaking, forbids any two-way interaction between distinct F-modules. Isolability only requires that information flow in at least one of the two directions be in some sense unnecessary. Thus, there may be a distinct syntax system even in the face of top-down semantic influence, so long as the mechanism that realizes the syntax system would still be able to perform its parsing task in the absence of this semantic information.
This strikes me as an interesting possibility, since interactions of this sort are probably commonplace. We know, for example, that the dorsal and the ventral streams in late vision are not neurally isolated; there are projections from each stream to the other. And even though this suggests (although it does not quite entail) informational interaction, the existence of dissociations indicates that each stream is able to perform some task without the aid of the other stream. Researchers who know neuroanatomy are usually quick to point out that there are afferent connections all over the brain, but many of these people are also making heavy use of the double dissociation paradigm to infer the existence of distinct systems.

Suppose that there is a semantics system, and if that system is removed, the substrate that computed syntax is still able to do so, but now computes a slightly different cognitive function (i.e., it no longer makes exactly the same input/output mappings as it did when it received feedback from the semantics system). Call the cognitive function computed when the semantics system is intact ‘S1’ and the function computed in the absence of semantic information ‘S2’. What we would have here (assuming that the substrate that computes S2 meets the other system criteria) is a host of overlapping systems: a syntax system which computes S2, a semantics system, and two importantly different systems that contain both of these. One of these two systems is a non-primitive system which contains syntax and semantics as subsystems -- the semantics system requires input from the S2 syntax system, and together they form a larger system. The other system is a (possibly primitive) syntax-and-semantics system, which computes S1, even though it may not deliver any syntactic representations as its output. We would have two distinct, though overlapping systems that perform slightly different parsing tasks, only one of which actually specializes in parsing, since the larger S1 system also performs the
task of semantic analysis.

The boxes and arrows scheme is not quite up to the task of distinguishing these systems. The main reason for this is that this scheme requires that overlap in mechanisms be reflected by overlap in boxes. Otherwise it would be impossible to represent the difference between nested systems (e.g., broader and narrower conceptions of the task) and redundant systems (i.e., two truly distinct systems that perform the same task). The boxes and arrows scheme does not give us different ways to enclose smaller boxes as subsystems on one hand and smaller boxes as comprising a primitive system on the other hand. And since the substrates that realize the syntax-and-semantics system also realize distinct syntax and semantics systems, these substrates have to be distinguished by drawing two small boxes. This, however, is only an indication of the limitations of the traditional box and arrow scheme, not in the present notions of systems or isolability.

The above example points to an important difference between isolability and informational encapsulation. Suppose that $p$ is a putative system which performs task $T$. If the removal of some other putative system, $o$, does not render $p$ incapable of performing $T$, then $p$'s ability to perform $T$ was isolable from the operation of $o$. However, $p$ need not have been informationally encapsulated from $o$; perhaps the information $p$ received from $o$ was redundant, or perhaps $p$ can still perform $T$ but is now forced to use a different algorithm -- to perform it in a different way -- due to the loss of information from $o$. Thus, informational encapsulation is a stronger notion than that of isolability: everything that is informationally encapsulated from something is isolable from it, but not every case of

---

10 I have not actually defined the notion of isolable from some particular substrate or mechanism, but it should be clear enough what such a notion would amount to: $S$ is isolable from $R$ with respect to $T$ if and only if $S$ performs $T$ and could do so even if $R$ did not compute any cognitive functions.
isolability is a case of informational encapsulation.

The distinction between various systems thus allows for top-down processing in a way that the distinction between various modules does not. If a higher-level system feeds into a lower-level system, then the lower-level system is not informationally encapsulated from the lower-level system. In such a case, the lower-level system would not be (F-) modular. Top-down processing is, however, consistent with a distinction between the higher-level system and the lower-level system, so long as the lower-level system is still able to compute a function without the aid of the higher-level system. No top-down processing can occur between two F-modules, but top-down processing can occur between systems, so long as it is inessential.

Just as isolability is a weaker relative of encapsulation, specialization is weaker than domain specificity. Because of the specialization constraint, all cognitive systems are going to be, in some relatively uninteresting sense, task specific, but the notion of tasks is far weaker than that of domains. Any set of outputs constitutes a task (individuated at the coarse grain), but not every set of outputs is supposed to constitute a domain. The notion of domains is supposed to be a substantive and robust notion in a way that the notion of tasks is not, and this actually poses something of a difficulty for the modularity view: it is far from obvious how to find a principled distinction between domains and non-domains.

Given certain possible cognitive systems, it is hard to tell whether or not they meet the domain specificity constraint. For example, is chess a domain? Is the past a domain? If there were a system that computed information about everything except the color of the shirt I am wearing right now, we would not want to call it domain specific merely on those
“By contrast with vision and language, there presumably is no well demarcated domain for the competence to go shopping for Xmas presents or to execute a spying mission behind enemy lines” (Segal, 1996). How are we to assess such a claim?

Certainly, tasks like remembering, or seeing, or comprehending language strike us intuitively as well demarcated in a way that spying and apprehending truths do not. But does this intuitive sense give us grounds for claiming that a mechanism that sees is domain specific, while one that apprehends truths is not? These sorts of questions matter for the F-modularity view, since domain specificity is supposed to be a necessary condition for F-modularity. Task specificity, on my view, does not help us to determine the presence of a cognitive system, since anything that computes any function at all is going to be task-specific: it will compute the task that it computes. When we say that there is a system for some task, we are saying that the system computes nothing else besides that task (this is why we would not say that there is a system for grandmother recognition). We are not saying anything about the intrinsic nature of that task. What seems like a gerrymandered task description to us might very well seem perfectly well demarcated to our brains. Thus the notion of cognitive systems developed here relies far less on prior intuitions about how to individuate tasks and domains.

11 Fodor himself admits that even “[c]entral systems may be domain specific in some sense” (p. 103), since there are probably things that we are incapable of representing. But this is not the sense of domain specificity that is at issue in discussions of modularity.

12 They matter for Fodor’s view, as well, since even though domain specificity is not taken to be a necessary condition, it is supposed to be a diagnostic feature.

13 As Rob Cummins has pointed out to me (in conversation), we might very well require some intuitive individuation of domains in order for us to count some task as cognitive. For example, we are more likely to consider a machine that can be interpreted as playing chess as performing a cognitive task than one that can only be interpreted as playing something like chess. And this relies on an intuitive understanding of what counts
2.2 A Special Case of OE Systems

I have argued that the notions of isolability and task specificity are weaker, and thus more inclusive, than the notions of informational encapsulation and domain specificity, respectively. Consequently, although all F-modules are cognitive systems, there might very well be cognitive systems that are not F-modules. Similarly, the etiological constraint offered in Chapter 4 and the notion of ratiocinative opacity are weaker than the criteria of characteristic ontogeny and the inaccessibility of interlevel representations, respectively.

The ontogenetic constraint is easily dealt with. F-modules are innately specified (Fodor, 1983; p. 37); OE systems are either innately specified or learned (or acquired via some interaction of innate constraints and learning). Innate specification is simply one kind of "normal" etiology.

In Chapter 4, I defined a ratiocinatively opaque faculty (or system) as one such that all or most of its outputs are cognitively spontaneous. A belief is cognitively spontaneous just in case it is not the result of an introspectible train of reasoning. The notions of inaccessibility of interlevel representations and ratiocinative opacity are obviously similar. However, I see no reason to think that every opaque system is going to be one whose interlevel representations are inaccessible. The mere fact that some interlevel representations are accessible does not mean that these representations serve as premises in an introspectible train of reasoning. In fact, there seem be systems which yield accessible interlevel representations that are not propositional in nature and thus could not even be the sorts of things that could be used as premises in reasoning.

as chess. Still, this involves much less reliance on our prior intuitions than modularity requires.
Now, I do not mean to be taking a position in the epistemological debate about whether non-propositional representations can contribute evidentially to the justification of beliefs. In fact, I tend to suspect that many things other than propositional representations might be able to confer justification (sensory states and the reliability of the relevant process are good contenders). The issue here just concerns the definition of cognitive spontaneity, a term which, as I intend it, does not imply total inaccessibility of interlevel representations, but only implies that whatever we do have access to does not, introspectively, appear to be a train of reasoning from premises to conclusions.

There is a potential misunderstanding lurking here, one which, I think, is due to a difference in Fodor's and my understanding of what perceptual systems are. Perceptual beliefs are paradigmatic instances of cognitively spontaneous beliefs, and perceptual systems are paradigmatic instances of opaque systems. Our perceptual beliefs are, inter alia, beliefs about physical objects in the world around us, beliefs which are not the result of an introspectible train of reasoning. However, if we take a perceptual system to be a transduction system that outputs beliefs about the identity, location, and so forth, of physical objects, it is clear that some of the interlevel representations computed by this system are accessible to consciousness. We have introspective access to perceptual states, and perceptual states are distinct from (and presumably causally prior to) perceptual beliefs.

The reason this might seem to be a problem is that perceptual systems are also supposed to be paradigmatic instances of modules. But how could this be if, as I have just claimed, they produce accessible interlevel representations? The answer to this, I think, can be found by returning to the notion of nested systems. The system that takes retinal stimulation as input and returns perceptual states as outputs is a nested subsystem of the
system that takes retinal stimulation as input and returns perceptual beliefs as outputs.\textsuperscript{14} The former might be F-modular even though the latter is not, and it is in this sense of 'perceptual system' (as the larger system) that perceptual systems can be OE systems without being F-modules. Under the other construal of 'perceptual system' just mentioned (i.e., as the subsystem), perceptual systems are possibly both OE systems and F-modules.\textsuperscript{15} My guess is that Fodor and proponents of F-modularity have the subsystem in mind. Epistemologists, however, are concerned primarily with beliefs and so, to the extent that they think about systems at all, would have the larger system in mind.

In sum, while all F-modules are going to be OE systems, there are likely to be many OE systems that are not F-modules. Innate specification is just one way of meeting the etiological constraint from Chapter 4. Similarly, while any system whose interlevel representations are all inaccessible will be ratiocinatively opaque, there might be opaque systems some of whose interlevel representations are accessible. It is in this sense that F-modules constitute a special case of OE systems.

3. Systems and Faculties

I promised earlier that I would try to explain the notion of faculties at work in the descriptive project by reference to the notion of cognitive systems. Developing an understanding of the folk psychological notion of faculties is important to the present

\textsuperscript{14} As was the case earlier, I do not mean to suggest that there is exactly one visual object recognition system. For the moment, however, it is harmless to pretend that there is as a simplifying assumption.

\textsuperscript{15} They might very well meet the opacity requirement only trivially, however, for these systems might not actually output beliefs at all. As I defined it earlier, a faculty (or system) is ratiocinatively opaque if and only if all or most of its doxastic outputs are cognitively spontaneous beliefs. Systems that have no doxastic outputs would meet this condition, although only trivially.
project for two reasons. First, the descriptive epistemological account offered in Chapter 4 relies heavily on the notion of faculties, and for that reason, the more we can say about what a faculty is, the more precise and predictive the resulting account will be, and the better situated it will be to handle objections. Secondly, since the normative epistemological theory developed in Chapter 4 utilized the reflective equilibrium approach, it is important that there be some continuity between the notion of faculties and the notion of systems. Given the reflective equilibrium approach to normative epistemology, if the notion of cognitive systems is just a precisification of the notion of faculties, then the mere fact that the latter appears in the descriptive account gives the former some right to appear in the normative account. If there is enough similarity between faculties and systems, then the opaque systems account from Chapter 4 will count as a faculty-oriented account. Absent such continuity, however, we would require some independent motivation for spelling out the normative epistemological account of privilege in terms of systems.

I think that the notion of cognitive systems developed above does bear the right sort of continuity to the notion of faculties, although these two notions differ in several important ways. A cognitive system, roughly, is something that is isolable and specializes in a unitary task. The notion of isolability is one that is essentially tied to the substrate that realizes the system. And so is the notion of unitariness, since it is defined in terms of isolability. I doubt that our commonsense understanding of faculties makes any appeal to the notion of an underlying substrate: in fact, it is likely that our commonsense understanding of the mind embodies a dualist ontology, and it is unclear whether such an ontology would even include a substrate in the sense of Section 1. In any case, I want to leave it an open question whether faculties, according to our commonsense conception, are in any interesting sense realized in anything. For this reason, I think that the notion of
faculties will have to rely on a more general notion than the technical notion of isolability. I will call this less specific, more vague and ontologically neutral notion "independence".

Before discussing independence in any detail, I want to point out another feature of our commonsense understanding of faculties that distinguishes it from the notion of systems developed earlier. A good deal of what is interesting about the cognitive scientific understanding of systems is that it assumes that there might be a great deal of unconscious mentality. The cognitive scientific boxology would contain far fewer boxes were it not for this possibility. There would be few, if any, subsystems within the visual object recognition system, for example, if there were no unconscious cognitive states computed as subtasks of the larger task of visual object recognition.

Our commonsense understanding of faculties does not seem to make this assumption about the possibility of unconscious cognitive states. Instead, our pretheoretical understanding of mentality is largely, if not exclusively, an understanding of conscious mentality.16 Now I suppose that it is possible that the Freudian doctrine of unconscious mental states has so deeply ingrained itself into our culture that some understanding of unconscious mentality has made its way into our folk psychology. This, however, is quite a different conception of unconscious mentality than the one operative in most of cognitive science. If and to the extent that our folk psychology recognizes unconscious mental states, these mental states are going to be things like beliefs and desires, rather than edge or geon representations. Our folk psychology does not seem to view perception or language comprehension or the like as resulting from a kind of

16 It is perhaps worth noting that Leibniz (1973) was the first philosopher to think that the notion of unconscious mentality was even coherent, and it took roughly another two hundred years for this to catch on. The view that all mentality is conscious (or at least potentially so) has been with us for a long time.
unconscious inference over representations to which we could never have direct introspective access. If our folk psychology posits unconscious mental states at all, it seems to posit only those mental states that are in principle available to consciousness. In this sense, I think our folk psychology has much in common with Searle's (1992) view.

Another important difference, then, between the notion of faculties and the notion of systems is that the latter notion assumes that a great deal of what goes on inside a system is unavailable to introspection. The former does not. The basic causal elements involved in a theory of systems will differ in this way from the basic causal elements involved in a theory of faculties. This puts certain limitations on an understanding of independence.

The conception of faculties as independent entities goes back at least to Plato (1941), who argued from conflicting mental states to the tripartite soul of the Republic. Plato, of course, was a dualist, which indicates that it is at least possible to view the mind as composed of parts without viewing the mind as being realized in some physical substrate. Now if the various parts of the mind can be in conflicting states at the same time, then it must be that what is happening in one part does not completely determine what is happening in another part. It is in this sense that the various parts of the mind are independent. There can, however, be interactions among the various parts of the mind; this is consistent with the sort of independence we see in Plato's conception of the mind.

The notion of independence here is, like the notion of isolability, a causal notion, but with three important differences. First, independence can be formulated in an ontologically neutral way (it is consistent with dualism and with physicalism), and secondly, the notion of independence does not explicitly depend on any counterfactuals about what would happen if some parts of the mind were removed. Thirdly, the notion of independence, unlike the notion of isolability, is not cashed out in terms of computing
functions. Just as isolability is a weaker notion than that of informational encapsulation, independence is weaker than isolability. Any two things that are isolable from each other will also be independent, but there might be independent things which are not isolable. Perhaps the only reason independence does not imply isolability is that it is so vague that it does not imply much of anything. However, since the notion of faculties is the one relevant to descriptive epistemology, and since I think that our pretheoretical understanding of faculties is itself vague, I am content to leave the notion of independence vague in this respect.

It seems that this notion of independence does have some bearing on our commonsense individuation of faculties. We are all aware that, for example, reasoning and perception sometimes conflict, that the deliverances of reasoning therefore do not completely determine the deliverances of perception, and vice versa. We even seem to utilize a sort of “folk neuropsychology”; we realize that hearing and vision are independent of each other, partly because we are aware of something like double dissociations. People can be blind without being deaf, and people can be deaf without being blind. Even such simple experiments as covering one’s ears or one’s eyes show that the two modalities are independent of each other in the relevant sense (i.e., in the sense that what is happening in one modality does not completely determine what is happening in another).

There is more to something’s being a faculty than its being independent. The unitariness constraint on systemhood is intended to rule out putative systems that are “too big”, and just as there is presumably no vision-and-olfaction system, our commonsense taxonomy of the mind’s parts does not recognize a vision-and-olfaction faculty, or even a
perception faculty. An analogue of the unitariness constraint seems to be in order for the notion of faculties. However, since there does not seem to be any commonsense notion of subfaculties, analogous to the cognitive scientific notion of subsystems, the analogue of the unitariness constraint is easier to formulate: no faculty contains any independent (proper) parts. The specialization constraint was trivial for systems (it only determined what we should call the system), and this constraint seems to transfer to the account of faculties virtually without emendation.

Altogether, we have four main features that distinguish the notion of systems from the notion of faculties. The notion of systems, unlike the notion of faculties, (a) is committed to a physicalist ontology, (b) understands tasks as computational functions, (c) views much of the causal interaction between and within systems as occurring below the level of consciousness, and (d) provides a more precise, counterfactual conception of independence (viz., that of isolability). Recall from Chapter 4 that by replacing OEF_D and OEF_D2 with OEF, we were correcting for mistaken folk psychological claims about what

17 Certainly we can talk about a “perceptual faculty”, but this terminology is merely intended to refer to any one of either vision, taste, touch, etc.

18 The definition of unitariness had it that a task is unitary with respect to substrate S if and only if S performs that task and no proper part of S is isolable with respect to any part of that task. Besides omitting the reference to the substrate, the analogue of the unitariness constraint offered above drops the reference to parts of tasks. This clause was needed in the definition of systems to distinguish non-primitive systems from collections of non-systems grouped together by gerrymandering tasks, since any non-primitive system is going to contain isolable parts. If the commonsense understanding of faculties does not include any understanding of subfaculties, then, in essence, it makes all faculties primitive, and for that reason, we do not need a clause that would distinguish non-primitive faculties from gerrymanded collections of faculties.

19 The only difference is that, where the specialization constraint for systems rests on a technical understanding of the term ‘task’ the specialization constraint for faculties will require (only) an intuitive understanding of tasks.
OE faculties there are. In the same way, the notion of systems developed here can be viewed largely as correcting for mistakes (or at least filling in lacunae)-in our commonsense understanding of the mind. The ontological commitment, the commitment to unconscious mental states, and the view of tasks as functions are simply the result of incorporating unintuitive facts about the mind (or at least we currently take them to be facts) to the conception of faculties.

The counterfactual account of isolability, on the other hand, involves more than just correction in light of new facts: there might be an alternative to the present conception of isolability, which would also count as a precisification of the notion of independence and yet would be incompatible with the one offered here. Still, the notion of isolability developed here strikes me as the most natural way to make independence more precise. Furthermore, this notion was developed with a particular methodology in mind (the neuropsychological methodology of double dissociations), and it seems that a similar methodology is involved in our commonsense individuation of faculties.

A faculty, then, is an independent part of the mind (in the sense that what is happening in the rest of the mind does not completely determine what is happening in it), which specializes in some task and contains no proper parts that perform any cognitive tasks. The notions of independence and of tasks are left fairly vague, here, since our pretheoretical understanding of faculties is similarly vague. Still, this says quite a bit more about faculties than was said in Chapter 4, which was merely that faculties are mental entities as opposed to mere dispositions. Having a specification of what faculties are is an essential element of the descriptive project, and the relation between faculties and systems is important for the normative project. If reflective equilibrium is to result in a theory according to which systems play the role that faculties do in the descriptive account, then
there needs to be a relatively clear sense in which the concept of systems is simply an improved version of the concept of faculties. I have argued that this is the case: the two rely on similar methodologies, and the concept of systems is deriveable from the concept of faculties, by making elements of the latter more precise and by taking into account certain facts about the nature of cognition more generally (e.g., that it is often unconscious, that it is a matter of computation over representations, etc.).

4. The No-OE-Faculty Objection

In Chapter 4, I left untouched a crucial counterexample to my descriptive epistemological account of the P/N distinction. According to this counterexample, it is possible for there to exist an agent who has no OE faculties. The creature might be doxastically and phenomenologically just like us (or sufficiently like us), but since it has no OE faculties, the present account of the P/N distinction would (incorrectly) imply that this creature has no privileged beliefs. An analogous objection could be leveled against the normative account by supposing that the creature has no opaque systems. OS would imply that this creature has no privileged beliefs, and this seems too strong. I postponed treatment of this case because I had not yet said enough about what faculties and systems are.

Now, I think, we are in a better position to see what such a scenario would have to entail. I will focus on the opacity constraint rather than the etiological constraint, because this is the one that survives the transition from descriptive epistemology to normative epistemology, and because I think that the case would no longer be even a putative counterexample if the creature in question were thought to acquire its faculties by some abnormal means. In such a case, our intuitions would be that the creature does not have
any privileged beliefs, and the present account would get this right. Our intuition that the creature does have privileged beliefs depends on the supposition that the creature is doxastically and phenomenologically sufficiently like us. Thus, I will assume that what is going on in this case is that the creature does have cognitively spontaneous beliefs, but because of an unusual cognitive architecture, it nonetheless does not have any faculties that count as opaque.

We can begin with a case that is clearly logically possible: a creature that has no opaque (and thus, no OE) systems. The hypothesis that our minds contain opaque systems is clearly a contingent and empirical one. There is no reason to think that there could not be a creature all of whose cognition is subserved by a single, general-purpose cognitive mechanism. This single cognitive system that is this creature's mind could fail to be an OE system, since it may not be ratiocinatively opaque, even though a large number of its beliefs may be cognitively spontaneous. OES and OS imply that this creature would have no privileged beliefs.

This general issue was already addressed in Chapter 4. In section 4, I noted that if further inquiry yields the discovery that we do not have any opaque systems, then OS would probably have to be abandoned. Even though a normative epistemological theory is not subject to refutation from contrary intuitions in the same way that a descriptive epistemological theory is, the reflective equilibrium approach implies that some results might be so counterintuitive, so unacceptable, that we would have to abandon the theory that produced them. If a theory, conjoined with the relevant psychological facts, were to imply that we do not have any privileged beliefs, then this would seem to be a result so counterintuitive as to require that we abandon the theory. The present theory only implies that it is possible that there be some creatures that have no privileged beliefs. This, to my
mind, at least, is far more acceptable. In fact, as we noticed in Chapter 4, it is possible that we are those creatures, and this mere possibility was not enough to reject the opaque systems account.

The overall plausibility of OS, and thus its acceptability as a normative epistemological account of the P/N distinction, is going to hinge on what the contingent facts turn out to be. In this sense, the logical possibility of creatures just like us but lacking any opaque systems, does not pose a threat to the present account. Nor would the physical possibility of such creatures (although I am far less certain that it even is physically possible for there to be creatures sufficiently like us but lacking opaque systems). It is only the actual existence of such creatures that would spell doom for OS, and as of yet, there is no reason to think that there actually are such creatures. Just as the final assessment of OS will have to wait until we know what opaque faculties we do, in fact, possess, it will also have to wait until we know whether there are creatures sufficiently similar to us in some epistemologically relevant respects but lacking opaque systems.

This general line of response, however, is of no use to the present descriptive epistemological account of the P/N distinction. The fact that Truetime does not really exist, for example, is of little comfort to one who defends simple reliabilism as a descriptive epistemological theory. Note, however, that the possibility of a creature that has no OE systems is not a counterexample to the descriptive epistemological account, since this account was cashed out in terms of OE faculties. While I admit that there might be creatures with no OE systems, I am not at all certain that there could be a creature that has no OE faculties and is still even remotely like us. The reason for this is that faculties are importantly different from systems, and this, in turn, is because independence differs in some significant respects from isolability.
An important difference between independence and isolability is that the cognitive scientific notion of isolability, unlike the pretheoretical notion of independence, allows for causal interactions at the unconscious level. We can understand the possibility of a creature with no opaque systems being phenomenologically and doxastically just like us, because we can understand the possibility of massively top-down or Quinean (in Fodor’s (1983) sense) processing occurring ubiquitously below the level of consciousness. There might be enough interaction that nothing would be isolable from anything else, but if this interaction occurred among unconscious mental states, the creature might operate just like us above the level of consciousness. If the creature operated differently above the level of consciousness, however, then it would not be doxastically and phenomenologically like us. The question at hand is whether a total lack of independence is possible for a creature that is nevertheless doxastically and phenomenologically like us (or enough like us that we still have the intuition that some of its beliefs are privileged).

It seems unlikely. Since the commonsense understanding of faculties does not recognize unconscious mental states, the only possible causal interactions would have to be among conscious mental states, in which case, the creature would not be phenomenologically like us. Since conscious mental states are the only ones the commonsense understanding recognizes, the sort of interaction implied by the lack of independence could hardly fail to be noticeable. If no parts of the mind were independent of other parts of the mind (in the sense of ‘independence’ described above), then there

---

20 Again, our folk psychology may include some beliefs about unconscious mental states, but if so, only about those mental states that are in principle available to consciousness. These mental states are not the ones that are at issue in the case of perception or memory, and so even if our folk psychology does include the claim that there are some (roughly Freudian) unconscious mental states, it will not make any difference to the present argument, since these sorts of states do not occur in our perceptual or memory faculties (or any other faculty whose outputs we intuitively take to be privileged).
could not be conflict between conscious mental states, since conflict, as mentioned in Section 3, implies independence.

A creature with a mind of this sort would be very different from us; such a creature could not, for example, have perception telling it that the stick is bent, while at the same time have reasoning telling it that the stick is (still) straight. A large part of our cognitive lives involves conflict, and not just in cases of perceptual illusion. We often seem to remember things that a little bit of reasoning tells us could not have happened; we are frequently surprised when we perceive something that memory tells us was not there a minute ago (and therefore, should not be there now). A creature who was incapable of being in conflicting mental states would be radically unlike us. It is not even clear that this creature would have anything that ought to count as perceptual beliefs or as memory beliefs. It might very well have beliefs, but none of these beliefs would bear a very strong resemblance to our perceptual beliefs. When we really get clear on what it would mean to be a creature with just one non-opaque faculty, I think that we tend to lose the intuition that such a creature must still have privileged beliefs.

Suppose we relax the assumption that the creature has just one, non-opaque faculty. There might be other ways in which a creature would fail to possess any opaque faculties. But is it conceivable that there exists a creature that has non-opaque perceptual faculties? Perhaps we can imagine a creature with non-opaque perceptual systems, but that is not the right question; the question has to be whether having non-opaque perceptual faculties is conceivable. Again, it is unclear to me that a creature so internally different from us ought to count as even having perceptual beliefs or perceptual faculties. Recall that a faculty is taken to be opaque just in case all or most of its doxastic outputs are cognitively spontaneous, and non-opaque otherwise. A non-opaque faculty is one a good number of
whose outputs are the result of an introspectible train of reasoning. Any argument for a perceptual faculty's being non-opaque seems to be a better argument for thinking that the faculty was not a perceptual faculty after all. But now there no longer seems to be anything particularly counterintuitive about the claim that this creature has no privileged beliefs. Which of this creature's beliefs does my descriptive epistemological account of the P/N distinction incorrectly classify as nonprivileged? Not its perceptual beliefs, for it is not even clear that this creature has perceptual beliefs.

5. Toward an Elaborative Epistemological Account

In this chapter, I have tried to clarify the natures of cognitive systems and cognitive faculties. Obviously, more effort was spent trying to get clear on the nature of cognitive systems, since the overarching epistemological project requires far more precision here than with the notion of cognitive faculties. In fact, the epistemological use to which I put the notion of faculties requires that this notion retain a certain vagueness.

I have tried to clarify the notion of systems by comparing it to the notions of F-modularity and of faculties. And I argued that the notion of systems is continuous with the notion of faculties and thus an apt candidate for the role it plays in the normative epistemological theory developed in Chapter 4. Finally, I used the clarified notions of cognitive systems and cognitive faculties to respond to one of the more pressing objections from Chapter 4.

Although I do not have the space here to give the elaborative epistemological project anything like the attention it deserves, I do want to say a few words about how an elaborative account might be derived from the above. Elaborative epistemology, recall, lies somewhere between descriptive and normative epistemology: it attempts to improve on our
commonsense epistemic concepts and principles but only by clarifying them, and not by replacing them. Any vagueness in the commonsense concepts or principles should be eliminated, but only under the condition that it can be eliminated without adding any new assumptions which contradict those implicit in the commonsense view. Exactly when clarification becomes replacement I do not profess to know, and this is one of the reasons I cannot give anything like a thorough defense of an elaborative epistemological proposal.

Before addressing the sorts of clarifications I think appropriate, we need to determine the conditions of formal adequacy for an elaborative account. Recall that the present approach to descriptive epistemology requires that a descriptive account be couched in terms of the judgments of attributors, and the normative account must exclude any reference to attributors. The attributor's (perhaps unconscious) epistemological convictions are made explicit in descriptive epistemology, and they are neither explicit nor implicit in normative epistemology. One way to keep the elaborative project in the middle ground between descriptive and normative epistemology is to require that the attributor's convictions are only implicit, but not explicit, in the account. We can do this by offering an account of the form 'a belief \( B \) is privileged if and only if \( p \)', where this account is offered, not as a theory of what makes a belief privileged, but as an elaboration on the commonsense convictions that drive our epistemological intuitions. Roughly, such an account would be tested not by seeing whether it correctly identified privileged beliefs, but by seeing whether it would be endorsed by an attributor who knew none of the relevant empirical facts other than those embedded in the knowledge structures responsible for her intuitive judgements of justifiedness. An elaborative epistemological proposal, then,

\[ \text{21 I say "roughly" here because this is intended only as a hint at how an elaborative epistemology should be pursued. The restriction to attributors who know nothing beyond what is encoded in the knowledge structures is intended to obviate the problems that arise} \]
would have the same formal structure as a normative proposal, but would be aimed at a
different target.

So how might an elaborative epistemological account set about improving on our
pretheoretical concepts and principles merely by clarifying them? I have argued that the
notion of systems improves on the notion of faculties in two different ways. First, it
brings in important but unintuitive (presumed) facts about the nature of the mind (e.g., that
much of our cognition is in principle unconscious, that minds are realized in physical
things). Secondly, the notion of systems clarifies the notions of independence and of tasks
by cashing them out in terms of counterfactuals and mathematical functions, respectively.
Now the former sort of improvement would be inappropriate for the purposes of
elaborative epistemology, since it relies on assumptions which are likely to contradict our
folk psychological convictions (although I would have no problem relying on those facts
that are consistent with our folk psychology). However, there is no reason not to help
ourselves to a counterfactual analysis of independence or an understanding of tasks as
mathematical functions in an elaborative epistemological theory. Augmenting the
descriptive epistemological account in this way will lead to a greater degree of precision,
without actually contravening anything included in the descriptive account.

Such augmentation will leave us with more precise conceptions of independence,
specialization, and the descriptive epistemological analogue of the unitariness constraint
(the latter two because they make reference to tasks). The resulting conception of
independence still differs from that of isolability, however, since, for example, isolability is

when the agent has explicit beliefs that conflict with the assumptions embedded in the
knowledge structures (Chapter 4, Section 2.1 and Section 2.3, pp. 138-40). It was
precisely this problem that mandated the explicit reference to the attributor and the
knowledge structures involved in her intuitive judgements of justifiedness earlier.
spelled out in terms of the realizing substrate, while independence is not. This, of course, is just a hint at how an elaborative epistemological account of the P/N distinction might proceed, but that is all it is intended to be. I think that even this rough sketch is enough to show, at least in a general way, how the descriptive, elaborative, and normative accounts of the P/N distinction would fit together. This exhausts what I currently have to say in defense of faculty-oriented accounts of the P/N distinction.
CHAPTER 6:
CONCLUSIONS AND CONSEQUENCES

So far I have been concentrating on a relatively small number of beliefs: my perceptual belief that there is something red in front of me or something round in front of me, and the beliefs that nominalism is true or that the mass of the top quark is 175 GeV or that it's going to snow a lot this winter. In addition, a lot was made of the beliefs of Norman, TrueTemp, and their alien counterparts. Although these are the sorts of beliefs that I used to argue for the P/N thesis and the etiological, faculty-oriented account of the P/N distinction, both of these (the P/N thesis and the faculty-oriented account) can do more for us than simply to help formulate an epistemological theory that will handle these sorts of cases. In this chapter, I want to point out some of the consequences of what has been done in the earlier chapters.

I begin by briefly rehearsing, in Section 1, the main arguments and positions developed in earlier chapters. In Section 2, I discuss the results of these findings for some major, general issues in epistemology: issues concerning the structure of epistemic justification, the internalism/externalism debate, and the debate about naturalized epistemology. Finally, in Section 3, I discuss the way in which the present theory might help to resolve some smaller, albeit still important, problems in epistemology.

1. Recapitulation

It will help to begin by summarizing what has been accomplished so far. I have been defending two main theses: (1) that there is a distinction between privileged and nonprivileged beliefs and that neither class is empty, and (2) that whether a particular belief
is privileged or nonprivileged is determined by the nature of the faculty that produced the belief. A privileged belief, again, is one that it is possible to be *prima facie* justified in holding even on the basis of an otherwise substandard argument (and no argument at all is just a limiting case of an otherwise substandard argument). A nonprivileged belief is one whose *prima facie* justification requires support from a generally adequate argument.

Foundationalists see perceptual beliefs and memory beliefs as either basic (POF) or directly inferable from basic beliefs (MSF). Other beliefs, according to both kinds of foundationalism, require generally adequate arguments for their *prima facie* justification. My belief that the Andromeda galaxy is 2.2 million light years away is neither basic, nor is it directly inferable from my belief that it seems to me that the Andromeda galaxy is 2.2 million light years away.

Coherentists see the *prima facie* justification of perceptual beliefs as arising from inferential relations to other beliefs, but the sorts of inferential relations adequate for the support of perceptual (and presumably memory) beliefs are not generally adequate. The sorts of arguments offered begin with the having of the belief and argue from that to the truth of the belief, and this does not work for many other kinds of beliefs. Furthermore, the kind of circularity involved in the justification of perceptual beliefs would be vicious if involved in the justification of many other beliefs.

Reliabilism was described as a version of POF, since it holds that some beliefs about physical objects can be *prima facie* justified in the absence of inferential relations from other beliefs. In fact, by claiming that reliability is sufficient for justification, simple reliabilism was seen to confer privileged status on far too many beliefs. This, I argued is what is really behind such famous counterexamples as the Norman and the Truetemp cases. Once we make the P/N distinction explicit, however, it becomes clear that these sorts of
cases are not counterexamples to reliabilism per se, but only to those versions of reliabilism that fail to countenance the existence of nonprivileged beliefs.

Throughout Chapters 2 and 3, I argued for the P/N thesis: the claim that some beliefs are privileged and some beliefs are nonprivileged. Any epistemological theory that fails to ascribe privileged status to any beliefs will result in skepticism, and any theory that fails to ascribe nonprivileged status to any beliefs will make justification too easy. It is not enough, however, to claim that there are some privileged and some nonprivileged beliefs: a thorough epistemological theory needs to ascribe privileged status to privileged beliefs and nonprivileged status to nonprivileged beliefs. In order to do this, we need an account of which beliefs are privileged and which beliefs are nonprivileged.

In formulating an account of the P/N distinction, I first distinguished between the descriptive and the normative epistemological projects. The former is aimed at capturing and systematizing our intuitions, while the latter attempts to improve on these intuitions. The descriptive epistemological account of the P/N distinction offered here is an etiological, faculty-oriented account. Our intuitions about privilege, I argued, are determined by our intuitions about the nature of the faculty that produced the belief. Glossing over a few complexities, we judge a belief to be privileged if and only if we judge it to be the output of a cognitive faculty that resembles our perceptual faculties and that has the right etiology. Somewhat more tentatively, I suggested that the relevant point of resemblance with perceptual faculties is ratiocinative opacity and that a faculty has the right etiology just in case it is either innate or learned.

Turning to normative epistemology, I suggested that we drop the etiological constraint (since it seems unprincipled) and that we replace the intuitive notion of faculties with the cognitive scientific notion of cognitive systems. A cognitive system is something
that is isolable and specializes in a unitary task. Our commonsense understanding of faculties, I argued, is understandably more vague than the scientific notion of systems, but the two do bear important relations to each other, enough so that the scientific notion can be viewed as simply a more precise and empirically informed version of the intuitive notion. Finally, I offered some brief suggestions as to how an elaborative epistemological account of the P/N distinction might proceed.

2. General Results

Of all the debates within epistemology, there are three that seem to have attracted the most attention in recent years, and which most clearly represent the “big picture” issues in epistemology. The first is the debate concerning the structure of epistemic justification: the debate between foundationalists, coherentists, and reliabilists. The second is the debate about internalism and externalism in epistemology. On some taxonomies, this second debate turns out to be just a generalization of the first debate, although the way that I defined foundationalism in Chapter 2 allows for both externalist (reliabilist) and internalist versions of foundationalism, so the debate is not merely about foundationalism or coherentism versus reliabilism, at least not on this taxonomy. The third debate is the one concerning naturalized epistemology, a debate about whether, and the extent to which, empirical concepts and findings are relevant to epistemology.

1 I defined coherentism in Chapter 2 as simply the denial of foundationalism. On this taxonomy, one could, in principle at least, defend a version of coherentism where ‘coherence’ is given an externalist explication.

2 A fourth debate that has occupied a central position in epistemology is that concerning the nature of knowledge. However, since most of this literature in recent years has been aimed at solving the Gettier problem, it represents less of a “big picture” issue than the others. Besides, I have nothing interesting to say about it, except insofar as I have something to say about justification, which is quite plausibly thought to be a necessary
The theory that I have developed here has something original to contribute to each of these debates. Or at least so I think. Each of these three topics is large enough that an exhaustive treatment of any of them would be impossible, and I will not even try. For the most part I will simply sketch what I take the contribution to be and leave it to the reader to locate the view with respect to existing views.

2.1 The Structure of Epistemic Justification

In Chapter 2, I floated a suggestion about the structure of epistemic justification that the P/N distinction made it possible to articulate.

(SEJ): Every justified belief is either privileged or ultimately derives its justification from a set of privileged beliefs.

I mentioned this in passing in Chapter 2 and did not even attempt to argue for it there, for the concern in that chapter was with the nature of the P/N distinction and its treatment by the major internalist theories. Although I think that this principle is intuitively plausible enough that some readers may not need much convincing, I do have something of an argument for it.

Recall the similarity between SEJ and the generic statement of foundationalism offered in Chapter 2. Indeed, the only difference is that the latter uses the term 'basic' where the former uses the term 'privileged'. Since all basic beliefs are privileged, foundationalism (including reliabilism) is simply a special case of SEJ, and therefore, any foundationalist is already committed to SEJ. However, SEJ is not simply a restatement of foundationalism, for there might be some beliefs that are privileged but nonetheless not basic. There might not even be any basic beliefs, in which case, if some beliefs are condition for knowledge.
justified, then foundationalism would be false, even though SEJ might still be true.

I argued in Chapter 2 that the plausible versions of coherentism are committed to the P/N thesis (i.e., the claim that there are privileged beliefs and nonprivileged beliefs). SEJ is thus consistent with a coherentist theory of justification, even if coherentism is defined as the denial of foundationalism. The P/N thesis, however, is a weaker claim than SEJ in a way, for the P/N thesis, per se, says nothing about the relation between the two kinds of beliefs. Could a coherentist theory of justification plausibly claim that there are justified nonprivileged beliefs that do not rely for their justification on privileged beliefs? This is what the denial of SEJ would require.

The argument for SEJ can be gotten by modifying the classical regress argument for foundationalism. The classical regress argument goes as follows:

Suppose some belief, $B$, is justified but not basic. Then it must be held on the basis of at least one other belief, $B^*$. But in order to confer justification on $B$, $B^*$ must be itself justified. If $B^*$ is not basic, then it must be held on the basis of at least one other belief, $B^{**}$. It is clear that if this regress of justification does not terminate in basic beliefs, then the support relations must be either infinite or circular. But an infinite regress of reasons cannot serve to justify a belief, and neither can a circular argument. So the regress must terminate in basic beliefs.

Now the way that coherentism avoids this conclusion is, of course, by claiming that circular arguments can, in fact, confer justification. According to coherentism, not all

---

3 There is also a way in which the P/N thesis is stronger, for SEJ says nothing about the existence of nonprivileged beliefs.
circularity is vicious.\(^4\)

I am willing to accept (at least for the present purposes) the claim that not all circular arguments are vicious. Some circular arguments might be adequate arguments, but this does not mean that circular arguments are generally adequate arguments. In fact, as I argued in Chapter 2, Section 2.3, circular arguments are not generally adequate (as is suggested by the fact that some circular arguments are vicious). This leads us to a new version of the regress argument:

Suppose some belief, \(B\), is justified but nonprivileged. Then it must be held on the basis of a generally adequate argument containing at least one other belief, \(B^*\). But in order to confer justification on \(B\), \(B^*\) must be itself justified. If \(B^*\) is nonprivileged, then it must be held on the basis of a generally adequate argument containing at least one other belief, \(B^{**}\). It is clear that if this regress of generally adequate arguments does not terminate in privileged beliefs, then the support relations must be either infinite or circular. But an infinite regress of reasons cannot serve as a generally adequate argument, and neither can a circular argument. So the regress must terminate in privileged beliefs.

This is all perfectly consistent with the central coherentist doctrine that some circular arguments do confer justification.

Thus, we have an argument for a view concerning the structure of epistemic justification, a view that is distinct from foundationalism, coherentism, and reliabilism, yet is consistent with any one of these. The fact that SEJ is in this sense not an actual

---

\(^4\) Another way to go is to endorse a kind of infinitism (Klein, unpublished) and insist that infinite regresses can justify beliefs. This view strikes me as implausible, but I do not have the space to discuss it in any detail here. In any case, the argument I offer below should apply, *mutatis mutandis*, to infinitism as well as to coherentism.
competitor to any of these theories should not be taken to show that SEJ is at all trivial or vacuous. In fact, there might very well be theories that would conflict with SEJ (I discussed some such possible theories in Chapter 2, Section 2.2); I just happen to think that there would be serious problems with any such theory.

2.2 Externalism

There are quite possibly as many different characterizations of internalism and externalism as there are epistemologists, and perhaps as many varieties of internalism as there are internalist epistemologists. A great deal has been written on the subject, and I will not even attempt to exhaustively survey the various characterizations and views. All I want to do here is to distinguish two different approaches to characterizing internalism and externalism and to argue that the foregoing argues for a kind of externalism, according to one of these approaches.

I argued in Chapter 4 that the difference between Norman and his alien counterpart indicates a kind of externalism with respect to the P/N distinction. It is easy to see how this also provides an argument for that same kind of externalism more generally, with respect to justification. The argument in Chapter 4 was that since Norman and Nyromoon are doxastically and phenomenologically identical, even though Nyromoon’s belief is privileged while Norman’s belief is nonprivileged, then one cannot always tell from the inside whether a given belief is privileged or nonprivileged. Obviously, there will be cases where an agent has some belief but lacks a generally adequate argument for it. In such cases, assuming that the agent has whatever kind of (otherwise) substandard argument would be sufficient for the prima facie justification of that belief if it were a privileged belief, the agent will not be able to tell from the inside whether that belief is prima facie justified or
not. If an agent cannot tell whether her belief is privileged or not, then she cannot tell whether her (otherwise) substandard argument is sufficient for the *prima facie* justification of that belief and thus cannot tell whether that belief is (*prima facie*) justified.

Let me distinguish two dimensions along which statements of internalism might differ. Internalism, in its roughest form, is the view that epistemic justification requires that something or other needs to be internally accessible in some sense. The first dimension, then, is that of what it is that needs to be accessible. The second dimension is the sense in which it needs to be accessible.

Let us focus for the moment on the first dimension. Compare the following two statements of internalism:

(I1): A belief is justified only if the factors that *serve to justify that belief* are internally accessible to the agent.

(I2): A belief is justified only if the factors that *determine whether that belief is justified* are internally accessible to the agent.

On the face of it, these two statements might seem to be almost identical. However, I think that there is an important difference.

Let us ignore, for the moment, what it means for something to be "internally accessible" to an agent, and let us simply suppose, for the sake of argument, that the privileged or nonprivileged status of a belief is not thus "internally accessible". This would imply that I2 is false, but it would not imply that I1 is false. I1 is a claim about what sorts of things can confer evidence; I2 is a claim about the nature of epistemic justification more generally. The privileged or nonprivileged status of a belief is a factor that (partially)

---

5 I concentrate on internalism, because this is how it is normally done and because we can then simply define externalism as the denial of internalism.
determines whether that belief is justified, but it is not a factor that serves to justify the belief. Being privileged is a metaepistemic property; it is one that determines the kind or amount of inferential support a belief needs in order to be justified, but it is not something that itself confers evidence on a belief.\(^6\)

What I have been concerned with all along is the difference between privileged and nonprivileged beliefs; I have been intentionally neutral on the question of what factors serve to justify beliefs (except to say that generally adequate arguments play some role). Specifically, I have been completely neutral as to what confers evidence on privileged beliefs and whether anything in addition to a generally adequate argument is required for the \textit{prima facie} justification of nonprivileged beliefs. It is possible to maintain this neutrality precisely because the P/N distinction occurs at a higher level than distinctions among potential justifying factors. Part of the point of even formulating the distinction between privileged and nonprivileged beliefs is that the factors that serve to justify a belief constitute a strict subset of the factors that determine whether that belief is justified.

However, even though I think that everything that I have said here is consistent with either II or its denial, this does not yet say anything about I2. Even though Norman’s belief and Nyrmoon’s belief receive evidential support from the same factors (whatever they may be), because Norman’s belief is nonprivileged and Nyrmoon’s is privileged, Nyrmoon might be justified while Norman is not. And the important point here is that neither agent can tell from the inside whether their belief is privileged or not and thus cannot tell from the inside whether it is \textit{(prima facie)} justified or not. So the theory here is

\(^6\) This is true even on a very liberal understanding of what counts as conferring evidence. In Chapter 1, I said that I was willing to count reliability, for example, as the sort of thing that might confer evidential support (though not \textit{inferential} support) on a belief.
inconsistent with at least those versions of I2 according to which a property is internally accessible if and only if the agent can tell, on the basis of mere reflection, whether she exemplifies that property (e.g., Chisholm, 1989).

This brings us to the second dimension along which statements of internalism might differ. I do not want to catalogue all the senses of ‘internally accessible’ that have been defended in print, and fortunately, the account(s) of the P/N distinction offered in Chapters 4 and 5 make this unnecessary.7 Recall that it is not mere cognitive spontaneity that makes a belief privileged, but being the output of an opaque faculty or system. Given any particular cognitively spontaneous belief, an agent cannot be sure whether the belief was the result of an opaque faculty or the result of a non-opaque faculty. We assume, of course, that we have a pretty good sense of which of our beliefs are perceptual beliefs and which are the result of, say, reasoning (even reasoning that is so automatic and fast that it might result in a cognitively spontaneous belief), but there is no reason why beliefs that were qualitatively identical to the beliefs we take to be perceptual could not be the result of a non-opaque system.

All this really shows, however, is that it is possible to be wrong about what faculty a belief comes from. It is possible that no internalist (besides, perhaps, Descartes) ever meant by ‘internal accessibility’ anything as strong as infallibility. A better argument for a kind of externalism here can be gotten at by distinguishing the descriptive account from the normative account. The descriptive account sketched in Chapter 4 was an etiological

---

7 Many of these senses would not be relevant to the present question anyhow because many of these are defined over justifying factors, rather than factors that determine wither a belief is justified. Consider, for example, Alston’s (1988) influential distinction between accessibility internalism, perspectival internalism, and consciousness internalism. Each of these offers a constraint on what sorts of things can confer justification on beliefs and thus serves as a way of fleshing out I1, but not I2.
account. Part of what makes us judge a belief to be privileged is our judging the faculty that produced it to have the right etiology. Now it should be fairly clear that the etiology of the faculty is not something to which we have internal access, under any reasonable construal of ‘internal access’. There is nothing internal to me that will distinguish at all between my perceptual faculties having existed for many years and their having existed for only a few minutes. It is not that we have only fallible internal access to etiological properties; we do not have internal access to these properties at all.

I was willing to drop the etiological constraint when it came to normative epistemology, but here the failure of internal access is even more obvious. The normative account of the P/N distinction made privilege a matter of being the output of an opaque system, in the cognitive scientific sense of ‘system’. In Chapter 5, I went on at some length about this notion of systems, and one thing that should have been abundantly clear is just how hard it is to determine exactly what systems there are. This is a project that requires painstaking empirical study. Again, it does not matter much what exactly might be meant by ‘internally accessible’; we have no more internal access to what cognitive systems we have than we do to the nature of the weak force in physics. Of course, if we do not have internal access to which systems we have, then we do not have internal access to whether a particular belief is the output of an opaque system. We would need to know not only which system produced the belief, but which other beliefs the system outputs and how many of these are cognitively spontaneous in order to know whether that system is opaque or not.

There is one more wrinkle to all of this that bears mentioning. Since the proposal here is not intended to express synonymous equivalence (Chapter 1, Section 1), it is possible to have beliefs about privilege without having beliefs about faculties, systems,
etiology, and the like. Thus, it is at least possible that our beliefs about privilege have an
evidential status not shared by our beliefs about faculties, systems, and the rest. It is
possible, for example, that we have an opaque system (with the proper etiology) which
delivers beliefs about which of our beliefs are justified. That is, it might be that our
intuitions about justifiedness are themselves the outputs of opaque systems. If that is the
case, then (some) beliefs about justifiedness will be privileged and hence justifiable on the
basis of an otherwise substandard argument or perhaps no argument at all. There might be
a sense in which this would count as internal access to the epistemic status of beliefs
without our having internal access to etiology, opacity, or any other properties that make
for this epistemic status.

This, however, would be an extremely strained sense of 'internal access'. Granted,
if these intuitions are privileged, then their (prima facie) justification does not necessarily
depend on inference from the standard sorts of external considerations. But if we claim
that any privileged belief is one whose truth is internally accessible, it is hard to see what
might not, in principle, be internally accessible. We might have privileged beliefs about
physical objects, but it would be odd, at the least, to claim that the external world is
therefore internally accessible. We might, for all we know, have opaque systems that
produce beliefs about the reliability of our belief-forming processes. This would make
such beliefs privileged, but if being privileged implies being internally accessible in the
sense required for internalism, then we have landed upon a quite uninteresting sense of
'internal accessibility'. Any understanding of internalism which does not exclude

8 Our intuitions about justifiedness are certainly cognitively spontaneous
(otherwise they would not count as intuitions). Claiming that they are the output of an
opaque system is a much stronger claim, since it is far from obvious which other beliefs are
output by the system that produces these judgments. Still, it is an interesting possibility,
and it would at least partially explain the evidential status of those intuitions.
reliabilism ought to be rejected, on grounds of triviality if nothing else.

2.3 Naturalized Epistemology

The term 'naturalized epistemology', like 'internalism', means many things to many people.9 Let me begin by distinguishing just two. A first kind of naturalism is a thesis about the ontological status of epistemic properties. According to this view, all epistemic properties are analyzable in terms of, or reduce to, or supervene on natural properties. Obviously, the etiological, faculty-oriented account of the P/N distinction developed here is consistent with this sort of naturalism, since it characterizes privileged and nonprivileged status in terms of natural properties. Just as obviously, however, it is consistent with a denial of this sort of naturalism, since it says nothing about many other epistemic properties, thus leaving open the possibility that some of these will not be thus reducible to natural properties. (The etiological, faculty-oriented account would, of course, be incompatible with a very strong anti-naturalism that would claim that no epistemic property is even extensionally equivalent to any set of natural properties.)

A second kind of naturalism -- and here is where I think most of the action is -- is a thesis about the relevance of empirical disciplines to epistemology. In its most general form, this second kind of naturalism claims that empirical scientific findings have something to say about epistemology. This is pretty vague, and intentionally so. In what follows I want to explain the extent to which, and the sense in which, the account of the P/N distinction developed here embodies a commitment to this second kind of naturalism. Henceforth, when I speak of "naturalism" or "naturalized epistemology" I will have this

9 For some taxonomies, see Kornblith, 1985; Goldman, 1994, forthcoming; Maffie, 1990.
second kind of naturalism in mind.

A generic sort of naturalism, then, claims that empirical disciplines are relevant to epistemology. There are three dimensions along which more precise statements of naturalism might vary. Which empirical disciplines (a) are relevant in what sense (b) to which aspects of epistemology (c)? We can refer to these three dimensions as the “discipline dimension”, the “relevance dimension”, and the “domain dimension”, or simply “discipline”, “relevance”, and “domain”, respectively. To take just one example, a notorious version of naturalism claims that behaviorist psychology (discipline) answers all the questions that can be answered (relevance) in that area of epistemology that is concerned with how epistemic agents manage to produce theories on the basis of sensory experience (domain) (Quine, 1969).\(^\text{10}\)

The second two dimensions are fairly self-explanatory. Empirical disciplines could be relevant to the study of knowledge, or of justification, or of the P/N distinction, or even (in principle, at least) narrower domains like the study of perceptual knowledge. If relevant at all, these disciplines could be relevant in the (very strong) sense of answering all the questions that can be answered, or perhaps only of providing constraints necessary for formulating an adequate epistemological theory, or perhaps only of making it easier to formulate an adequate epistemological theory.

It is impossible to say much about the discipline dimension without introducing some technical terminology, and it would be helpful here to clarify just what I take some of the various candidate disciplines to be. Most naturalists agree that the scientific discipline most likely to be relevant to epistemology is cognitive science, but there are several finer-

\(^{10}\) This, I think, is a fairly charitable reading of Quine, one of whose mistakes seems to have been that of thinking that this aspect of epistemology is all there is to epistemology.
grained distinctions among disciplines within cognitive science. I use the following terms in what I take to be a standard way, but there may not be canonical usages for all of these.

By 'cognitive psychology', I mean that branch of psychology that uses behavioral data to make inferences about the underlying causes of cognition. 'Behavior' is used very loosely here, in such a way that "behavior" can range anywhere from verbal reports to reaction times to bar pressing. Cognitive psychological theories generally focus on the content and the structure of the representations that would predict the observed behavior. As such, cognitive psychological theories are virtually always couched in semantic-level (i.e., cognitive-level) terms. Contrast this with (straight) neuroscience, which is a purely realization-level discipline. Neuroscience is the study of the nervous system (including, but not limited to, the brain); as such, it is a branch of biology, and has little to do with psychology.

Neuropsychology is the discipline which applies cognitive psychological methodology to impaired populations in an effort to determine which brain regions are involved in the performance of which tasks. Neuropsychology, therefore, unlike cognitive psychology or straight neuroscience, involves an attempt to bridge the cognitive and the realization levels, although in a limited way. A more general attempt to bridge these levels is cognitive neuroscience, which subsumes the findings and methodologies of cognitive psychology, neuropsychology and neuroscience, along with a few others, like the various functional imaging techniques, in order to bridge the cognitive and realization levels. Cognitive neuroscientific theories, however, are not limited to simply describing which brain areas are involved in which tasks; they also (sometimes) attempt to explain how various representational states are realized in the brain.

Finally, evolutionary psychology is that branch of psychology which attempts to
predict and explain behavior and cognitive-level structure on the basis of considerations of selective advantage. Testing the evolutionary predictions often requires the application of cognitive psychological or social psychological methodology. Much of the work done in evolutionary psychology (especially that which relies on social psychological methodology) is aimed at non-cognitive phenomena, like mate preference, and is thus not a very likely candidate for contributing to epistemology. However, a good deal of evolutionary psychology is concerned with cognitive phenomena and so might have some bearing on epistemological issues.

Now my main concern here has been with the descriptive epistemological project, and I think that the general approach here is one that is conducive to a sort of naturalism. The descriptive epistemological project, as conceived in Chapter 4, is one of describing the knowledge structures that are responsible for our intuitive judgements of justifiedness. These intuitive judgements are the data from which the character of the relevant knowledge structures are inferred. Both the data (intuitions) and the target (the nature of the knowledge structures) make cognitive psychology an appropriate tool for studying this phenomenon; in fact, one of the primary functions of cognitive psychology is to use behavioral data (e.g., verbal reports, reaction times, etc.) to make inferences about the structure and content of the knowledge structures operative in the performance of some cognitive task. Cognitive psychology (discipline) is helpful in determining (relevance) the nature of the knowledge structures that underlie our intuitive judgements of privilege (domain). So far, this does not differ much from the sort of naturalism Goldman (1992a) urges with respect to descriptive epistemology, except that the domain differs slightly, since my concern is with the P/N distinction and his is with justification.

However, I think that there are grounds for an even stronger claim. In Chapter 4, I
mentioned that to provide a thorough defense of any specific descriptive epistemological account of the P/N distinction, we would need some independent evidence for thinking that we possess the sort of knowledge structures invoked by OEF and its variants. In addition, I pointed out that the task of characterizing the knowledge structures would be easier if we had some assurance that the folk psychological principles involved in ascriptions of privilege were the same ones that are responsible for our intuitive judgements about others' mental states. It is hard to see how we might come across such independent evidence about the relevant knowledge structures without relying on empirical psychology. Cognitive psychology, therefore, might not be merely helpful; it might be practically impossible to provide a genuinely adequate descriptive epistemological account of the P/N distinction without help from cognitive psychology. This is a stronger claim about relevance that the one above.

Furthermore, if having adequate evidence for a descriptive epistemological theory requires having adequate evidence about whether the knowledge structures involved in our intuitive judgments about privilege are distinct from the knowledge structures involved in other kinds of intuitive judgments, then it looks like cognitive neuroscience will be able to play some role. And this is a stronger claim about the discipline dimension than the one above.11

The normative epistemological theory developed in Chapters 4 and 5 suggests a version of naturalism which is very strong along the relevance dimension. If being the output of an opaque system is necessary and sufficient for being privileged, then we simply

11 I do not propose to have any strict criteria about what makes a claim about discipline stronger than another, but the idea is intuitive enough. The weakest claim, which still might possibly count as naturalistic is that commonsense has something to tell us about epistemology. A much stronger claim would be that physics or astronomy has something to tell us about epistemology.
will not be able to determine which beliefs are privileged without the aid of cognitive science. This is not as strong a take on relevance as Quine’s view, for I do not think that all the questions that can be answered can be answered by the empirical cognitive scientific disciplines. The normative epistemological theory developed earlier is (if correct) part of the answer to the question of which beliefs are privileged and which beliefs are nonprivileged, and it would be implausible to claim that such a theory could have been developed without doing philosophy. Still we are left with a strong claim about relevance.

But which disciplines are relevant? Here is where I think that the present normative epistemological account is most distinctive. Many naturalistic epistemologists have granted that cognitive psychology and/or evolutionary psychology are relevant, and perhaps even indispensable. But to my knowledge, at least, no one has claimed that cognitive neuroscience would be indispensable. I think that cognitive neuroscience is indispensable to this project; while the various cognitive psychological methodologies might indicate a few system boundaries, we are not (as a matter of practical necessity) going to be able to provide an exhaustive list of cognitive systems without relying on cognitive neuroscience. And if that is the case, then we will not be able to partition the set of beliefs into privileged and nonprivileged beliefs without reliance on the deliverances of cognitive neuroscience.

This, I think, is a surprising result. Many epistemologists are willing to claim that psychology might be indispensable to epistemology, but as far as I can tell, I am the only one who thinks that brain lesioning, imaging studies, and/or studies of the neuroanatomical structure of the brain might be even helpful, let alone indispensable. But I have argued that such studies are practically necessary for discovering what systems there are. Now I am willing to concede that we might be able to get by without doing lesion studies and imaging studies and neuroanatomical studies, but it seems that we will have to do at least one of
these, and this still leaves us with a strong claim about discipline (and the relevance of that discipline): cognitive neuroscience is indispensable in determining which beliefs are privileged and which beliefs are nonprivileged.

Finally, the way in which these disciplines are relevant diverges from what we usually see in defenses of naturalism. Any epistemology that includes a causal/historical account of the basing relation should hold that cognitive science is relevant in that it can inform us about the actual causes of our beliefs (e.g., Goldman, 1986). And reliabilists frequently hold that cognitive psychology, and perhaps even evolutionary psychology, can help to tell us whether certain processes are reliable (e.g., Stich 1984 for a somewhat skeptical application of this). Finally, some naturalists claim that empirical disciplines are relevant to epistemology in that they inform us of what humans are capable of, or at least what tasks are more and less difficult for humans, and to the extent that 'ought' implies 'can', then empirical findings will place limits on what normative principles it is reasonable to demand of normal human epistemic agents (e.g., Goldman, 1992a). The present account of the P/N distinction suggests a different way in which empirical disciplines might be relevant: by informing us about our functional architecture, in particular, which cognitive systems we have and which of these are opaque. As far as I know, no other naturalists have argued for this sort of relevance.12

3. Particular Results

The three "big picture" debates just mentioned are usually construed as debates about the nature of knowledge or justification in general. However, there are also a

12 Goldman (1992a, p. 165), for example, explicitly denies that he has cognitive systems in mind when he claims that empirical science can tell us about the processes that result in our beliefs. His topic there, of course, is justification, rather than privilege.
number of less central, though still important, epistemological debates about various particular ways of knowing or being justified. Epistemologists have traditionally distinguished between perception and inference, for example, and between *a priori* and *a posteriori* justification and knowledge. I think that the P/N distinction and the account of it developed here can shed some light on such traditional epistemological issues. Furthermore, although these get quite a bit less press, there are a number of issues in what we might call "special interest epistemology", e.g., what, if anything, justifies our moral beliefs, or our religious beliefs? Here, too, I think that the material from earlier chapters might provide some answers.

### 3.1 Perception and Inference

In Chapter 2, I relied on an intuitive understanding of which beliefs count as perceptual beliefs, and I made a point of sticking to uncontroversial examples, examples like 'there is something round in front of me' or 'there is something red in front of me'. The reason this mattered is that there are controversial examples. A longstanding epistemological problem concerns the question of where we should draw the line between perception and inference. Descartes (1984), for example, claimed that when I look out a window and claim to see people on the street beneath me, all I genuinely *see* are coats and hats and *infer* from this that there are people inside them. Similarly, Berkeley (1948) claimed that one does not actually *hear* the coach driving past, but only hears a sound that leads one to infer the existence of the coach.\(^{13}\) Reid (1967, p. 184), on the other hand,

---

\(^{13}\) There are, to be sure, differences between these two claims, and not just in virtue of the fact that one is about seeing and the other is about hearing. Descartes' claim in the *Meditations* might (though there is little textual evidence one way or the other) be read as anticipating G. E. Moore's (1959) claim that all we ever really see are the *surfaces* of objects. Berkeley does not seem to have anything like this in mind.
claims that there is a sense of ‘perception’ (viz., acquired perception) according to which we really do perceive these sorts of things (in fact, Reid explicitly mentions Berkeley’s coach example).

This general sort of debate has carried on into the present century. Churchland (1979, 1985) thinks that some of us can taste the chemical composition of wine and see wavelength distributions. Brandom (1994) goes so far as to claim that a scientist can observe a mu meson moving through a vapor chamber. We do not (or do not merely) infer these things from what we perceive, but we actually perceive them. Other philosophers are somewhat less explicit, but seem to have a much more restricted understanding of what we perceive and correspondingly assign more credit to inference. Consider, for example, Chisholm’s account of the justification of our beliefs about who we are seeing:

In reply to the question, “What is your justification for counting it as evident that it is Mr. Smith whom you see?” a reasonable man ... would say ... something like this: “(It is evident that:) Mr. Smith is a tall man with dark glasses; I see such a man; no one else satisfying that description would be in this room now ... etc.” (1982a, p. 81; italics and last ellipsis in original).

Chisholm does not explicitly claim that such beliefs are not perceptual beliefs, but the sort of inference we see here is certainly very different from the sort we see for beliefs like ‘there is something blue in front of me’. If my (visually-based) belief that Mr. Smith is in front of me does not count as a perceptual belief, then it is hard to see why my (visually-based) belief that there is a proton in front of me should count as perceptual. How are we to assess such claims? How and where do we draw the line between perception and
inference?\footnote{The authors who have offered analyses of perception (e.g., Chisholm, 1957; Dretske, 1969; Jackson, 1977) tend not to say much about this question, because the analyses are typically given in terms of looks or appearances. However, the answer to the question of whether it can look as if this is Smith in front of me is no more obvious (to me, at least) than the answer to the question of whether I can have the perceptual belief that this is Smith in front of me. (Similar issues arose in Chapter 2, Section 1.2.)}

One terminological ambiguity needs to be resolved before pursuing this in any detail. The aforementioned authors often talk in terms of perceiving \( x \), e.g., hearing the coach or tasting the chemical composition of the wine. Such claims can be given either a referentially transparent or a referentially opaque reading. Many authors prefer to use the ‘perceives that’ locution, as this is always opaque to substitution of coreferential terms and thus avoids the ambiguity just mentioned. One can see that the cat is in the house without seeing that Susan’s favorite pet is in the house, even though the cat is Susan’s favorite pet. There is a sense, however, in which one cannot see the cat without seeing Susan’s favorite pet, since the cat and Susan’s favorite pet are the same thing. Clearly, when Berkeley claims that we do not hear the coach, he means this claim in the opaque sense just mentioned. His claim is intended to be uncontroversial (even though it is not), and so he must mean that we do not hear \textit{that} the coach is passing by, but instead hear something else from which we infer that the coach is passing by.\footnote{It is interesting to note that when Reid expresses his disagreement with Berkeley, he does so explicitly in terms of ‘that’ clauses.} Churchland’s claims, on the other hand, are supposed to be controversial, and so he, too, must have the opaque sense in mind. That is, he must be claiming that we can perceive \textit{that} the wine has such-and-such chemical composition. One way to avoid such ambiguity is to speak in terms of perceptual beliefs, rather than perceiving, since beliefs are always opaque. This is what I will do for the most part; however, when I do talk about “perception” -- as opposed to inference, or...}
what have you -- I will intend it in this opaque sense.

I will suppose without argument that the aforementioned authors are making epistemological claims about perception, and not (purely) psychological claims. Thus, the question of which beliefs count as perceptual beliefs is intended here as an epistemological question. I have framed this question in terms of the perception/inference distinction, but the first thing we can note here is that, since the question at hand is an epistemological question, this distinction turns out to be ill-formed.

One of the important results of drawing the P/N distinction in the way that I do is that some beliefs might very well be both perceptual and inferential, at least in the epistemological senses of these terms. The fact that perceptual beliefs are privileged does not show that they are not inferentially justified, since it is perfectly consistent with a belief’s being privileged that it is inferentially justified (albeit only via an otherwise substandard argument). The distinction that emerges here is not the distinction between perception and inference, but between perception and non-perception. Still, the P/N distinction does show us what is epistemologically distinctive about perception (even though there may be plenty of non-perceptual beliefs that have the same property), and this is that perceptual beliefs are privileged. At least the clear, paradigmatic examples of perceptual beliefs are, and if we are interested in an epistemological distinction between perception and non-perception, then this is a feature that ought to play a significant role.

Consider, for example, a theory of perceptual belief that simply farmed the problem out to psychology. One could claim that a belief is a perceptual belief if and only if (the true, finished) psychology says it is. My guess is that, for the purposes of psychology, the most important distinction between perceptual beliefs and other beliefs will be that the former involve modality-specific representations, where the latter do not (or perhaps the
former do in a manner in which the latter do not). This, at least, seems to be a promising approach to distinguishing perceptual representations from non-perceptual ones and perhaps it could be extended to apply to beliefs as well. As plausible as it may be for psychological purposes, however, it has no apparent epistemological significance. It has nothing to say about the one thing that makes perception a distinctive way of knowing and being justified: that perception results in privileged beliefs.

A better approach to drawing the line between perceptual beliefs and non-perceptual beliefs, at least for epistemological purposes, is to build the privileged character of perceptual beliefs into the definition of perceptual beliefs. We could say, then, that a belief is a perceptual belief if and only if it is the output of an opaque perceptual system. The resulting concept of perceptual belief is a mixed psychological/epistemological concept, and this seems right. We would let cognitive science tell us what perceptual systems are but not what perceptual beliefs are, of the concept of perceptual beliefs, unlike the concept of perceptual systems, is intended to be an epistemologically significant concept. I started here with the clear cases of perceptual beliefs and found that they were privileged beliefs. If perceptual beliefs are to constitute an epistemic kind, then at least one of the necessary conditions for being a perceptual belief should be an epistemic condition, and being privileged is the most salient epistemological property that the central cases of perceptual

---

16 Something like this seems to be in the air again, after a pretty long hiatus. Bever (1992), for example, claims that one of the reasons why there have to be perceptual modules is that informational encapsulation results from the fact that perception uses a kind of representational scheme distinct from, and largely incompatible with, that in use by the central, higher cognitive, systems. The idea of different, incompatible representational schemes for perception and conception was, I think, first suggested by Kant (1965); this incompatibility was the problem that the Schematism was intended to solve.

17 I am forwarding this proposal in the language of the normative epistemological theory, although it is easy to see how a descriptive epistemological account of perceptual beliefs would go.
beliefs seem to have.

Of course, it will be an empirical matter which beliefs are perceptual beliefs in the present sense of ‘perceptual beliefs’, although the present sense of ‘perceptual beliefs’ is one that is sensitive to what is epistemologically distinctive about perception. On the present proposal, it is not the intrinsic properties of a belief, but the etiological properties of the belief that make it a perceptual belief or not. And this is because it is these very same properties that make it privileged or not. This definition of perceptual beliefs explains why it is that agents with very different cognitive organization have different sets of propositions that can be the objects of perceptual beliefs for them. I do not have any opaque perceptual faculties that produce the belief that it is 104 degrees outside, but some creatures might, and this is why I never have the perceptual belief that it is 104 degrees outside, although some creatures might.

Now it might be objected that there is something circular about the above argument. In Chapter 2, I used paradigmatic instances of perceptual beliefs in order to argue for the claim that perceptual beliefs are privileged, and now I turn around and define perceptual beliefs in terms of that privilege. This approach strikes me as unobjectionable, however; it is no different from a standard particularist methodology (Chisholm, 1982b) in philosophy. According to this methodology, we start with a number of objects that instantiate the property in question, and only then answer the question of what it is for an object to have that property. For example, we take a number of beliefs that we intuitively think are justified, and use these as the basis for producing a theory of justified belief. This theory, once developed, can be used to determine the status of the borderline cases, the cases we were not sure of initially and thus did not use in formulating the theory. There seems to be nothing wrong with this general approach.
The present proposal about which beliefs count as perceptual beliefs bears certain important affinities with Reid's view. Reid's notion of acquired perception allows learning to play a role in perception. Recognizing the sound of the coach as the sound of the coach is something that requires experience, but this does not mean that we cannot hear that a coach is driving past. Similarly, I have allowed that opaque faculties and systems (even OE faculties) may result from learning. However, the proposal I am offering is probably going to be more restricted than Reid's (and almost certainly more restrictive than Churchland's and Brandom's), for merely recognizing the sound of a coach as the sound of the coach (or a certain wavelength distribution as a certain wavelength distribution, etc.) is not generally going to suffice to make the resulting belief the output of a perceptual system. Perceptual learning takes time and repetition (Gibson, 1969), and the mere fact that I now associate this sound with this object does not at all mean that my auditory system now produces beliefs about that object. There may very well be propositions that will never be the objects of perceptual belief, despite any amount of perceptual training. Our visual systems may learn to produce as outputs beliefs about the sex of chicks, but it is likely that no amount of training will enable them to produce beliefs about wavelengths or protons.

With this theory of perceptual beliefs in hand, it is possible to offer a new theory of perception in the vein of Chisholm (1957), Dretske (1969), and Jackson (1977). We might, for example, claim that S perceives that a is F if and only if S has the perceptual belief that a is F, and this belief meets whatever requirements are necessary for it to count as knowledge. Such a theory would, unlike the aforementioned theories, provide a unified account of perceiving, perceptual belief, and the justification of perceptual belief. Furthermore, it might hold a special appeal for those theorists who eschew an overtly
causal theory of knowledge, as the standard analyses of perceiving tend to wear their causal requirements on their sleeve. One interesting difference between this and the traditional analyses is that the present theory does not include any requirements about looks or appearances. One reason this is appealing is that, as mentioned earlier, it is not clear what counts as looks and appearances. In addition, the present account allows for the possibility of perception without any phenomenology at all, and depending on how ‘looks’ and ‘appearances’ are cashed out, the other theories might not. This strikes me as an attractive result. Robots which lack qualia ought still to be able to perceive things; it ought to be possible to have zombies that are not blind.

3.2 The A Priori

I also think that the P/N distinction might be able to do some of the epistemological work that the notion of a priori justification is supposed to do. Traditionally, the notions of a priori knowledge and justification were supposed to explain how we could have apodictic certainty about necessary truths. There has been a great deal of debate about the nature of a priori knowledge and justification since then, and one mainstream conception of the a priori is now considerably weaker than it used to be. Some theorists doubt that there even is such a thing as a priori justification. For these reasons and a few others, I think we might be better off simply replacing talk about a priori justification with talk about

---

18 I say “overtly causal” because many, if not all, theories of knowledge will in fact require a causal connection for perceptual knowledge, even though they do not explicitly contain a causal condition. Defeasibility theories of knowledge, for example, required that there not be any true proposition which would serve to make the belief in question no longer justified. There is no overt causal requirement here, but any time an agent’s perceptual belief is not caused by the object it is about, it would fail to count as knowledge, for this truth about the causal relation between the object and the belief would defeat the agent’s justification for that belief.
privileged beliefs.

Very roughly, a belief is justified *a priori* just in case it does not derive its justification from experience. Traditionally, this was taken to have far-reaching epistemological consequences. According to Leibniz (1973), for example, the only way one can know something prior to experience is if its denial is a contradiction. And since contradictions are impossible, anything we can know *a priori* must be true in all possible worlds. Furthermore, since it is impossible to conceive of a contradiction, it is impossible to be *a priori* justified in believing something that is false: *a priori* implies infallible. Locke (1975) and Hume (1978) saw *a priori* intuition as the immediate comparison of ideas, and as such it is the simplest cognitive act we are capable of. Thus, *a priori* intuition would have the highest degree of certainty. An explanation of how *a priori* justification was obtained would explain how we could have certainty about necessary truths, and in particular, about necessary truths *qua* necessary truths. 19

None of these important epistemological consequences of the *a priori*, however, follow directly from the characterization of *a priori* justification as justification independent of or prior to experience. Each relies on some independent assumption about how we could have *a priori* justification. If *a priori* justification does not all depend on the principle of non-contradiction, for example, then the above argument for necessity is undermined. If *a priori* intuition is not merely a matter of comparing two ideas, then it might be possible to have *a priori* justification that is less than certain. And so forth.

Because these epistemological consequences rely on additional assumptions in this way, it is possible to hold a much weaker conception of *a priori* justification. In fact,

---

19 It is one thing to know that 1+ 2 = 3 and another to know that *necessarily*, 1+ 2 = 3. An explanation of *a priori* knowledge would explain both of these, not just the former.
several recent authors, who believe in *a priori* justification, have been willing to deny that
(i) all *a priori* justified beliefs are infallible (Burge, 1993; BonJour, 1998; Goldman, forthcoming),
(ii) all *a priori* justified beliefs are certain (Burge, 1993; BonJour, 1998; Goldman, forthcoming),
(iii) all *a priori* justified beliefs are necessary (Kripke, 1972), (iv) all *a priori* justified beliefs are true (Burge, 1993),
(v) all *a priori* justified beliefs are immune to undermining from beliefs held on the basis of experience (Burge, 1993; BonJour, 1998; Goldman, forthcoming).

It is this conception of *a priori* justification -- justification independent of experience, but which is possibly fallible, revisable, etc. -- which I will have in mind in what follows, except where I explicitly contrast it with the stronger, traditional conception.

Thus sanitized, however, it is hard to see what interesting epistemological consequences the notion of *a priori* justification retains. In fact, I am inclined to doubt that this understanding of *a priori* justification buys us anything that the notion of privileged belief does not already get us.21 To make this argument, it is important to distinguish between *a priori* beliefs and *a priori* inferences (see BonJour, 1994; 1998). By "an *a priori* belief" I just mean a belief that is justified *a priori*, in the sense described above. An *a priori* inference, or argument, following BonJour (1994, 1998), is one that provides a reason for thinking that the conclusion follows from the premises, in a way that is

---

20 This is not, of course, to say that all contemporary authors have such a restricted view of the *a priori*. Katz (1998), for example, maintains something much more like the traditional view.

21 Note that this is a much more modest conclusion that some authors have drawn. Kitcher (1983) and Quine (1953), for example, have taken (rational) revisability to indicate that there is no such thing as *a priori* justification. I am not claiming that by allowing for revisability, fallibility, and the like, the sanitized conception of the *a priori* fails to refer or is in any way incoherent; I am merely claiming that it loses much of its epistemological significance.
independent of experience (BonJour, 1998). According to BonJour (1994), this can only be the case if the argument is one whose corresponding conditional is justified \textit{a priori}. Thus, the notion of \textit{a priori} inference or argument simply reduces to the notion of \textit{a priori} belief. For this reason I will concentrate on \textit{a priori} beliefs in what follows, largely ignoring the issue of \textit{a priori} inference.

Locke and Hume distinguished \textit{a priori} intuition from \textit{a priori} demonstration, and this is an important distinction.22 “[S]ometimes the Mind perceives the Agreement or Disagreement of two \textit{Ideas} immediately by themselves, without the intervention of any other: And this, I think, we may call \textit{intuitive Knowledge}. For in this, the Mind is at no pains of proving or examining, but perceives the Truth, as the Eye doth light” (Locke 1975, pp. 530-1). Demonstrative knowledge, on the other hand, is knowledge of the agreement or disagreement of ideas that is mediated by the interposition of other ideas, where each of the intermediate connections is itself intuitive. \textit{A priori} intuition, then, results in a belief that is justified both independently of experience and without any inferential support from other beliefs. (I will also call this resulting \textit{belief} an “\textit{a priori} intuition”.). \textit{A priori} demonstration results in beliefs that are justified independently of experience, but which are justified only via inferential support which consists of the chaining of intuitions.

It is important to notice here that it is really the \textit{a priori} intuitions that are doing all

\footnotesize
\begin{itemize}
\item[22] In order to avoid getting embroiled in exegetical issues, I am adding the modifier ‘\textit{a priori}’ here at the risk of being redundant. Although a fairly standard interpretation of the Lockean and Humean conceptions of intuition and demonstration is as a distinction between two kinds of \textit{a priori} knowledge, this is not entirely clear. Locke, for example, thinks that we can have intuitive knowledge of our own existence (1975, p. 552), even though it is hard to reconcile this with his own definition of intuition. Locke, in the \textit{Essay} (1975), and Hume, in the \textit{Treatise} (1978), both avoid the terms ‘\textit{a priori}’ and ‘\textit{a posteriori}’. Although Hume does use them in the first \textit{Enquiry} (1970), he does not use them to distinguish knowledge (intuition and demonstration) from probability.
\end{itemize}
the work. Although the notion of *a priori* demonstration is significantly narrower than BonJour's conception of *a priori* inference (the latter can result in beliefs which are not themselves *a priori* justified), they are similar in that *a priori* demonstration reduces to *a priori* intuition. The justification of *a priori* intuitions explains the justification of *a priori* demonstrations, because all demonstration is a number of intuitions, strung together.

If we take seriously the contemporary, sanitized understanding of the *a priori*, however, it becomes clear that all of the epistemological work done by the notion of *a priori* intuition is already done by the notion of privileged beliefs. Some *a priori* beliefs (viz., the intuitive ones) are privileged beliefs, e.g., 'a thing cannot be red all over and green all over at the same time', '1 + 1 = 2', '(p & (p ∨ q)) → q'. These beliefs, intuitively, are justifiable in the absence of any generally adequate argument. Other *a priori* beliefs, e.g., the Pythagorean theorem, are nonprivileged beliefs, but it is easy to see how such beliefs can be justified by chaining together a number of *a priori* beliefs, such that each step in the argument is one whose corresponding conditional is a privileged (*a priori*) belief.

Claiming that these beliefs are privileged does not yet explain why they are justified, because the notion of privilege developed here is neutral with respect to what makes privileged beliefs justified. But when embedded in a larger epistemological theory, the notion of privilege does explain the justification of these beliefs. To take a simple example, one might claim that all privileged beliefs that result from a reliable process are *prima facie* justified. The fact that S's belief that 1+1 = 2 is privileged and the result of a reliable process thus explains why S's belief is *prima facie* justified. Does adding that S's belief is *a priori* explain anything more? The belief's being privileged does not explain why it is certain, infallible, *ultima facie* justified, or the like, but on the sanitized understanding of the *a priori* described above, the belief's being *a priori* does not explain these either,
since a belief might be *a priori* and not have any of these properties.

In fact, claiming that a belief is *a priori* does not, by itself, even explain why the belief is *prima facie* justified, for the same reason that claiming the belief is privileged does not. To say that a belief is *a priori* is just to say that it derives its justification from reason alone, or some other non-experiential source. This does not tell us how or why pure reason confers justification on beliefs. To explain even the *prima facie* justification of *a priori* beliefs we need a larger epistemology that contains some epistemic principle linking pure reason with justifiedness. As is the case with explanations of the justification of privileged beliefs, it is likely that, for example, the reliabilist and the internalist foundationalist will offer different principles here. Thus, an account of privilege explains what is epistemically significant about *a priori* beliefs as well as an account of the *a priori* would.

One objection that might arise here is that there is one important thing that an account of the *a priori* would explain which an account of privilege would not. *A priori* intuitions are such that to consider (and understand) them is to be justified in holding them; privileged beliefs are not like this. More accurately, one might claim that there is a certain class of *propositions*, such that if one considers any one of those propositions, then that person is (*prima facie*) justified in believing that proposition. And an account of the *a priori*, unlike an account of privilege, would explain this.

This, however, is simply false. An account of the *a priori, per se*, would not explain why some propositions are such that to consider them is to be justified in believing them, for not all *a priori* beliefs are intuitions. Surely it is possible to consider the

---

23 This sort of claim seems to be contrary to the spirit of the sanitized understanding of *a prioricity* that I have been discussing, but I will address this objection anyhow.
Pythagorean theorem, for example, without being justified in believing it. To explain why considering some proposition implies being justified in holding that proposition, we would need an account of intuition. Such an account would have to tell us why we can intuit that $1 + 1 = 2$ but not that $17472 \div 56 = 312$. That is, it would have to tell us which propositions are susceptible to intuition and which propositions are not.

The account of privilege developed earlier is well-poised to explain why beliefs like '1 + 1 = 2' are privileged and beliefs like '17472 \div 56 = 312' are not. The simple answer is that we have opaque systems that produce beliefs of the former sort and not of the latter sort.\textsuperscript{24} Would this explain the connection between considering $p$ and being justified in believing $p$? Yes, in a way. Some of our opaque systems might very well be such that, for some set of propositions, they produce as output belief in that proposition, whenever the agent entertains the proposition. This would mean that whenever we consider one of those propositions, the corresponding belief is privileged for us.

This is weaker than the claim that whenever we consider them, they are justified (because some privileged beliefs are not justified, or even \textit{prima facie} justified), but this seems perfectly appropriate. A reliabilist would not want to say that my belief that $1 + 1 = 2$ is actually \textit{justified} whenever I consider it, unless it is the result of a reliable process whenever I consider it. Some coherentists will want to deny that this same belief is justified for me, unless I also believe that I am trustworthy in what I accept. To the extent that being privileged is weaker than being justified, this is an attractive feature of the current proposal. The claim that some belief is privileged is also weaker than the claim that it is

\textsuperscript{24} Although she does not claim that they are the outputs of opaque systems, Wynn (1992) has argued that rudimentary mathematical abilities (e.g. the ability to add and subtract very small numbers) are innate and that more sophisticated mathematical abilities involve the use of some other mechanism. It would not be surprising if the rudimentary abilities were subserved by opaque systems.
intuitive. But again, this seems appropriate. Coherentists, at least, should rather talk about privileged beliefs than about intuitions, since an intuition is simply a particular kind of basic belief.

A defender of the *a priori* might object that there is too much contingency floating around here. An obvious implication of my claim that the notion of privilege does all the epistemological work done by the sanitized notion of the *a priori* is that which propositions are such that, considering them is nearly sufficient for being *prima facie* justified in holding them, will depend on my cognitive construction, and this is a contingent fact about me, rather than a logical feature of the propositions. But we already knew that contingent facts about cognitive construction were relevant to this issue. Even on the traditional view, surely it is only because of how we are built that we cannot simply intuit the truth of the Pythagorean theorem. There might be some highly advanced race, the members of which are able to intuit the truth of the Pythagorean theorem. Similarly, it is because of how we are built that we can intuit the fact that $1+1=2$ and did not need Russell and Whitehead to prove it to us. Other creatures might be very different. Kant (1965) thought that some knowledge of Euclidean geometry was necessary (for any rational agent) for the very possibility of (outer) sensation. But it is easy to imagine rational agents for whom Riemannian geometry is the "form of outer intuition" and for whom Euclidean geometry seems foreign and bizarre. Which propositions can be intuited by whom depends on the cognitive structure of the agent in question, on both my view and the traditional view. And so my view is capable of explaining even the fact (if it is a fact) that there are some propositions such that to understand them is to be justified in holding them.

Thus, I think that an account of privilege would explain the justification of *a priori* intuitions as well as a sanitized account of the *a priori* would, and since demonstrative
belief is just belief based on intuitions strung together, an account of privilege would explain the justification of these as well. And, importantly, it can do so while avoiding the controversy that surrounds claims about \textit{a priori} justification and knowledge. Several authors (notably, Quine, 1953; Kitcher, 1980) have expressed worries about whether there \textit{is} any such thing as \textit{a priori} justification. We already have to countenance privileged beliefs, as I argued much earlier, and this can account for the justification of alleged \textit{a priori} intuitions without positing any additional mode of justification.

In addition to the controversy surrounding the existence of the \textit{a priori}, there is some controversy surrounding the nature of the \textit{a priori}. One reason for identifying \textit{a prioricity} with analyticity was that, until Quine anyhow, it seemed to be a relatively straightforward question whether a proposition is analytic or not, and this would yield a simple criterion for determining whether any given belief is \textit{a priori}. Claims about the synthetic \textit{a priori}, however, are much harder to assess, and this creates a number of dialectical difficulties to which the P/N distinction is far less susceptible. The claim that murder is wrong, for example, might be (synthetic) \textit{a priori} and it might not be \textit{a priori} at all. How are we to tell which it is? Suppose I claim to have \textit{a priori} justification for my belief that it will rain tomorrow. I concede that this belief is neither necessary, nor certain, nor analytic, nor immune to refutation from experience, but insist that none of these features is essential to \textit{a priori} justification. How would one argue against such a claim? An opponent could insist that claims about \textit{a prioricity} are themselves \textit{a priori}, but it is obvious that this will not get us very far. If, on the other hand, I claimed that the belief that it will rain tomorrow is privileged, then there is a good deal that could be said in response. We have a way of resolving this dispute, by appealing to the account of the P/N distinction offered earlier. We might have to defer to empirical science to resolve the dispute, but at
One of the primary uses of the notion of a priori intuitions has been to explain how we could be justified in holding certain beliefs for which we could not produce a (generally adequate) argument. The present account of privileged beliefs does just this and, I think, does it better, since it provides a relatively clear criterion for what counts as a privileged belief and avoids the controversies surrounding a priori justification. It is for these reasons that I think we might be better off simply replacing talk about a priori justification with talk about privileged belief.

3.3 Additional Possibilities

The present theory concerning the P/N distinction might help to answer a number of additional, even more specific, epistemological questions concerning certain domains of justified belief. The questions I have in mind here are questions like "(how) are our moral beliefs justified?" and not questions like "(why) do we intuitively think that some of our moral beliefs are justified?" Thus, these questions are intended as normative epistemological questions, rather than descriptive epistemological questions. I will focus on the normative epistemological account from earlier, although it will be easy to see how the descriptive epistemological account would help to answer the former sort of question as well. Again, the exact details of OS are less important to me than the general faculty-oriented account, although I will use these details for the sake of illustration.

OS claimed that a belief is privileged if and only if it is the output of an opaque cognitive system, and I developed a detailed account of what a cognitive system is in Chapter 5. Putting these together, we have a relatively precise criterion for a belief’s being privileged, even though it is one that a normal agent could not even begin to apply in the
course of an average day. It becomes a difficult, but tractable, empirical question what systems there are and which of these are opaque.

Empirical investigation might reveal that opaque systems are responsible for at least some of our moral beliefs. If this is right, then some of our moral beliefs are privileged. Of course, I think that this sort of claim is one that will require cognitive scientific investigation. One traditional view in moral epistemology has it that propositions like 'murder is wrong' are self-evident, and therefore basic. An important difference between this sort of view and the present suggestion is that I am claiming merely that these beliefs might be privileged. This does not mean that they are necessary, certain, infallible, basic, or even true. All it means is that they are *prima facie* justifiable in the absence of a generally adequate argument.

The same holds for religious belief. It is an empirical possibility that we have an opaque system that produces belief in God. Plantinga (1993a), for example, believes in a *sensus divinitatis*, a faculty that yields religious knowledge. Now Plantinga does not tell us whether this faculty counts as a *system* in the present sense, whether it is opaque in the present sense, or even why he thinks such a thing exists. Still, if there is an opaque system that produces belief in God, then belief in God would be privileged. Again, this does not imply that God exists or even that anyone is justified in believing in God (since we all might be in possession of undermining evidence), but it would serve as a good starting point for someone who is interested in the epistemology of religious belief.

The theory of privileged and nonprivileged beliefs might also contribute to discussions in feminist epistemology. There are almost certainly *some* psychological differences between men and women, and it strikes me as a genuine possibility that men and women may have slightly different opaque systems. (This is especially plausible in
light of the role that learning might play in the development of cognitive systems.) I do not predict anything so dramatic as the difference between us and Nyrmoon or Vipertemp, but something more like the difference between non-expert chicken sexers and expert chicken sexers would not be surprising. Again, this is something that would require empirical investigation, but it should not be dismissed out of hand.

One final topic I want to mention here is testimony. Some recent authors (e.g., Burge, 1993; Coady, 1992; Sosa, 1994) have argued that there is something special about the epistemic status of the beliefs that we hold on the basis of testimony. One argument for this view (Coady, 1992; Sosa, 1994) is that we simply do not have sufficient inductive evidence to support our belief in the general reliability of testimony. These authors are reluctant to say anything very explicit about what follows from this, but what would seem to follow is that our testimonial beliefs, if justified, are basic (since we do not have adequate nondemonstrative support for them, and it is clear that no demonstrative argument would be appropriate). This, of course, seems too strong, which is perhaps why the proponents of the argument just mentioned avoid drawing this conclusion explicitly.25

The P/N distinction would allow us to say more precisely what is and is not distinctive about testimony, provided that we are convinced by the argument that claims that we do not have adequate inductive evidence for the general reliability of testimony.26 Suppose that we cannot produce a generally adequate argument for any of our testimonial

25 Instead, they say things like “testimony seems as basic a source of knowledge as the traditional perception, memory, introspection, and inference” (Sosa, 1994; p. 67; italics added).

26 As it turns out, I do not think that we should be convinced by this argument (Lyons, 1997), but I explore this possibility here, because I suspect that some readers might be convinced.
beliefs. Would it follow that these beliefs are basic? No; it would only follow that these beliefs are privileged, since they still might require the support of an otherwise substandard argument. Would it follow from the privileged character of testimonial beliefs that testimony is a special source of epistemic justification? Yes and no.

The P/N distinction is intended to clarify the sense in which certain beliefs are distinctive (privileged beliefs, unlike nonprivileged beliefs, are justifiable on the basis of an otherwise substandard argument) and at the same time, to show the sense in which these beliefs are not distinctive (any privileged belief is epistemologically pretty much like any other privileged belief). If testimonial beliefs are privileged, then they are distinctive in the sense that they are unlike our philosophical and scientific beliefs, but this provides no reason to think that they are interestingly different from our perceptual beliefs, our memory beliefs, and possibly many other kinds of privileged beliefs.

What makes these special topics in epistemology (like the nature of the justification of our moral beliefs, religious beliefs, or our testimonial beliefs) interesting is that these are beliefs for which it is notoriously difficult to provide a generally adequate argument. At the same time, however, (some) epistemologists are convinced that (some of) these beliefs are justified, anyhow. The P/N distinction makes it easy to see how this could be true: if these beliefs are privileged, then their justification does not require the support of a generally adequate argument. At the same time, however, the account of the P/N distinction developed earlier prevents a facile appeal to the notion of privilege in an attempt to vindicate whatever beliefs one might hope to vindicate. It is a substantive and (as I hope Chapter 5 indicates) fairly well-constrained, empirical question which beliefs are the outputs of what kinds of faculties. It is in this way that I think that the P/N distinction and the account of the distinction developed here can have consequences for epistemology beyond that of
capturing intuitions about the cases that motivated the theory in the first place.
REFERENCES


Academic Publishers.


