VARIETIES OF NATURALIZED EPISTEMOLOGY: CRITICISMS AND ALTERNATIVES

Benjamin John Bayer, M.A.
Department of Philosophy
University of Illinois at Urbana-Champaign, 2007
Jonathan Waskan, Adviser

Naturalized epistemology—the recent attempt to transform the theory of knowledge into a branch of natural science—is often criticized for dispensing with the distinctively philosophical content of epistemology. In this dissertation, I argue that epistemologists are correct to reject naturalism, but that new arguments are needed to show why this is so. I establish my thesis first by evaluating two prominent varieties of naturalism—optimistic and pessimistic—and then by offering a proposal for how a new version of non-naturalistic epistemology must move forward. Optimistic naturalism attempts to use scientific methods to give positive answers to traditional epistemological questions. Epistemologists, for example, are urged to draw on psychology and evolutionary biology in order to show our beliefs are justified. I argue that this project fails. First, the naturalist's thesis that theory is underdetermined by evidence poses difficulties for the optimist's attempt to show that our beliefs *are* justified, even according to naturalized standards. Second, while critics usually contest naturalists' logical right to use the concept of normative *justification*, I suggest that a deeper problem is with the naturalists' use of the concept of *belief*. Naturalistic philosophy of mind, while perhaps acceptable for other purposes, does not deliver a concept of "belief" consistent with the constraints and needs of naturalized epistemology. Pessimistic naturalism—Quine's project—takes it for granted that "belief" is problematic and logical justification elusive, and instead offers a pragmatic account of the development of our theory of the world. This project, while deeply unsatisfactory to the traditional epistemologist, also faces the challenge of privileging scientific discourse over other pragmatically successful modes of discourse. Whatever its merits, we can undermine its motivation by challenging the underdetermination thesis it rests on. We can do this by appealing to facts about scientific practice that undermine the conception of confirmation driving the thesis, by appealing to other facts about scientific practice, and by challenging some philosophical preconceptions, in order to make room for a new brand of inductivist foundationalism.

**VARIETIES OF NATURALIZED EPISTEMOLOGY:
CRITICISMS AND ALTERNATIVES**

BY

BENJAMIN JOHN BAYER

B.A., Lawrence University, 1998
M.A., University of Illinois at Urbana-Champaign, 2000

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Philosophy
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2007

Urbana, Illinois

ABSTRACT

Naturalized epistemology—the recent attempt to transform the theory of knowledge into a branch of natural science—is often criticized for dispensing with the distinctively philosophical content of epistemology. In this dissertation, I argue that epistemologists are correct to reject naturalism, but that new arguments are needed to show why this is so. I establish my thesis first by evaluating two prominent varieties of naturalism—optimistic and pessimistic—and then by offering a proposal for how a new version of non-naturalistic epistemology must move forward. Optimistic naturalism attempts to use scientific methods to give positive answers to traditional epistemological questions. Epistemologists, for example, are urged to draw on psychology and evolutionary biology in order to show our beliefs are justified. I argue that this project fails. First, the naturalist's thesis that theory is underdetermined by evidence poses difficulties for the optimist's attempt to show that our beliefs *are* justified, even according to naturalized standards. Second, while critics usually contest naturalists' logical right to use the concept of normative *justification*, I suggest that a deeper problem is with the naturalists' use of the concept of *belief*. Naturalistic philosophy of mind, while perhaps acceptable for other purposes, does not deliver a concept of "belief" consistent with the constraints and needs of naturalized epistemology. Pessimistic naturalism—Quine's project—takes it for granted that "belief" is problematic and logical justification elusive, and instead offers a pragmatic account of the development of our theory of the world. This project, while deeply unsatisfactory to the traditional epistemologist, also faces the challenge of privileging scientific discourse over other pragmatically successful modes of discourse. Whatever its merits, we can undermine its motivation by challenging the underdetermination thesis it rests on. We can do this by appealing to facts about scientific practice that undermine the conception of confirmation driving the thesis, by appealing to other facts about scientific practice, and by challenging some philosophical preconceptions, in order to make room for a new brand of inductivist foundationalism.

ACKNOWLEDGEMENTS

Writing my dissertation took far longer than it should have, but not because *this* dissertation took so long. The problem was the three or four previous dissertation topics, none of which ever got off the ground. So the first person I would like to thank is Gary Ebbs, who saw me through all of these many topic changes, even after he left UIUC for another school. His knowledge and interest in Quine is what inspired and enabled the present project. I would also like to thank Jonathan Waskan in particular, for providing lots of useful and detailed feedback, especially during the last year in Gary's absence. Thanks are extended also to the other members of the committee, Arthur Melnick and Patrick Maher, particularly for their patience in waiting for the drafts which were submitted in sudden fits and starts. I would also like to thank my friend Greg Salmieri, for countless hours of philosophic conversation, much of which informed the most crucial topic-area decisions I had to make. For endless moral support, I would also like to thank Greg and also my friend Marc Baer, my parents David and Mary Ann Bayer, and my wonderful girlfriend Melissa McEwen.

TABLE OF CONTENTS

CHAPTER 1

CONCEPTUAL AND DOCTRINAL PROJECTS IN NATURALIZED EPISTEMOLOGY

There is a widespread belief among intellectuals that the domain of philosophy shrinks as the domain of the special sciences expands, and that, someday, science might swallow up philosophy entirely. Some philosophers—philosophical naturalists—believe that this day may have already arrived. Naturalists hold that philosophy does share or should share the basic concepts and methodologies of natural science.

To determine whether the naturalists are right, one useful approach is to examine proposals for naturalistic (or naturalized) *epistemology,* the recent attempt to transform theory of knowledge into a branch of natural science. In Western philosophy, epistemology has long been considered one of the most distinctively philosophic subjects. If even it can be naturalized, the days of philosophy as an autonomous discipline could be numbered.

Traditional epistemologists object to naturalized epistemology on the grounds that it eliminates the distinctively philosophical content of epistemology. In this dissertation, I argue that traditional epistemologists are correct to reject naturalism, but that new arguments are needed to show why they are correct. I establish my thesis first by critiquing two prominent versions of naturalism—which I call "optimistic" and "pessimistic"—and then by offering a proposal for how a renewed non-naturalistic epistemology must move forward.

Before I can outline how I plan on critiquing these two varieties of naturalism, I need to provide some important background exposition. In this introductory chapter, I will describe just what naturalized epistemology is supposed to be, in particular what it means for epistemology to "share the same concepts and methodologies of natural science." It turns out that apart from sharing this very generic credo, advocates of naturalized epistemology have deep differences over what it means for epistemology to be continuous with science. I will show how these different recognized approaches fit into my categorization of "optimistic" and "pessimistic" naturalized epistemology.

Having surveyed different conceptions of the naturalist's project, I will then describe one of the most prominent objections to it: the charge that naturalism unnecessarily eliminates the normativity of epistemology. I will briefly sketch the responses naturalists typically offer. With this as a background, I will describe how my own distinctive critique of naturalized epistemology compares to this traditional objection, and outline the course this objection will take through the rest of my dissertation.

The variety of conceptual and doctrinal projects

In his influential essay "Epistemology Naturalized" (1969a), W.V. Quine draws a distinction between "conceptual" and "doctrinal" projects in the traditional epistemology to which his naturalism is presented as an alternative. I find it useful to invoke this distinction to explain distinct but related projects within naturalized epistemology itself. Even though Quine critiques the manner in which traditional epistemology attempts to base its doctrinal project on its conceptual one, I find that many versions of naturalism follow the same pattern. (Whether or not Quine's naturalism does the same is somewhat more obscure.)

Quine begins by discussing the conceptual project in mathematics, which he compares to a similar project in epistemology. This project is concerned with "clarifying concepts by defining them, some in terms of others" (69). The doctrinal project is concerned with "establishing laws by proving them, some on the basis of others." (69–79). Quine then notes that the two projects are closely connected:

> For, if you define all the concepts by use of some favored subset of them, you thereby show how to translate all theorems into these favored terms. The clearer these terms are, the likelier it is that the truths couched in them will be obviously true, or derivable from obvious truths. (70)

In epistemology, the doctrinal project attempts to explain how we might justify "our knowledge of the truths of nature in sensory terms" (71), whereas the conceptual project aids by defining the terms of that knowledge. Famously, Quine argues that the traditional epistemological project of translating concepts of physical objects into the language of sense data had to fail, because of his indeterminacy of translation thesis. This failure, combined with the failure of traditional foundationalist proposals, spelled the death of traditional doctrinal projects in epistemology—not only the classical empiricist attempt to justify scientific knowledge by reference to the senses, but even the modern empiricist attempt to legitimize scientific discourse by "demarcation."

No naturalized epistemologist is interested in traditional epistemology's reductivist conceptual project or foundationalist doctrinal project. However the conceptual-doctrinal distinction is still at play for many naturalists, although at a higher, *meta*-epistemic level. While naturalized epistemologists no longer concern themselves with translating the content of empirical knowledge for the sake of justifying it, many are still concerned with analyzing or in some way defining the *concept* of "knowledge" itself, in order to answer the *doctrinal* question of whether and to what extent we have any knowledge in the sense provided by that definition.

In what follows, I first classify naturalized epistemologists according to their "optimistic" and "pessimistic" answers to the doctrinal question. The optimism and pessimism here is in relation to the traditional goals of epistemology, which I myself share. "Optimists" affirm that we can show human beliefs to be justified, by applying some naturalized conceptual project. "Pessimists" deny this, but would not consider themselves to be pessimists, because they urge that epistemology adopt new goals.

Optimistic naturalized epistemologists are united in the conviction that the empirical methodology of natural science can somehow show our beliefs to be justified, but there is a variety of views about what this methodology amounts to. Not surprisingly, every major semantic theory of the twentieth century—analytic, two-factor, natural kinds—has been applied to the project of understanding the reference of the concept of knowledge. I will, therefore, classify subvarieties of optimistic naturalism according to the semantic theories they rely upon.

Having presented these optimistic projects, I will turn to the pessimists. The first of these is Michael Williams (1996), who offers a "deflationary" approach to the concept of knowledge, which focuses on the use of the *term* "knowledge," rather than its reference in the world. The most prominent pessimist, however, is Quine himself. Though Quine would, in some moods, speak of human knowledge, the concept of "knowledge" does not figure prominently as a technical concept in his naturalized epistemology. Quine's behaviorism generally rendered the epistemologist's reference to subjects' internal cognitive states to be of largely passing concern. As we will see in later chapters, Quine's deep commitment to the principles of naturalism not only caused him to distance himself from the very idea of a conceptual project, but from many of the philosophical mechanisms used by epistemologists (naturalistic or otherwise) to engage in this project.

Optimistic naturalized epistemology

Analytic naturalized epistemology

The first putatively naturalist epistemology worth discussing engages in a meta-epistemic conceptual project with deep ties to traditional epistemology. This approach seeks to offer genuine conceptual analyses of epistemic concepts such as "knowledge" and "justification," but hopes to analyze these concepts into more basic concepts that are naturalistically respectable. This approach is exemplified in the epistemology of Alvin Goldman. Goldman's early views sought to analyze the normative language of "justification," for example, into the purely descriptive terms such as "'believes that', 'is true', 'causes', 'it is necessary that', 'implies', 'is deducible from', 'is probable'." These latter terms are "(purely) doxastic, metaphysical, modal, semantic, or syntactic expressions" and therefore neither epistemic nor normative (Goldman 1979, 2).

Examples of this analytic approach to the naturalistic conceptual project originally gained prominence as responses to the Gettier problem. One challenge of that problem was to identify a condition for knowledge that would explain why justified true beliefs that were merely "accidentally" true did not count as knowledge. A natural solution was to individuate knowledge by the *causal* origin of the belief. David Armstrong's account, for example, treats knowledge as a kind of reliable indicator, like a thermometer, in which “reliability” is understood as a lawlike connection between beliefs and facts (1973). Robert Nozick’s (1981) theory speaks of knowledge as “tracking” the truth, and analyzes “reliability” in counterfactual terms: a true belief counts as knowledge just in case the following holds: if it was true, it would be believed, but not otherwise. Goldman’s own (1986) version of reliabilism holds that a belief is justified just in case it results from a reliable belief-forming process, one that yields a greater percentage of truths than falsehoods, and counts as knowledge if it is both true and discriminates the truth from “relevant alternative” possibilities.

In the next chapter we shall examine whether the mere non-epistemic or non-normative status of doxastic, metaphysical, modal or semantic concepts is sufficient to guarantee their status as naturalistic. For the time being, however, the more interesting question is whether the approach of conceptual analysis itself is consistent with naturalism. Recognizing that the armchair approach of analysis has long been rejected by naturalists, Goldman urges that any adequate epistemology seems to “involve, or presuppose, analyses (or ‘accounts’) of key epistemic terms like ‘knowledge’ and ‘justification’” (Goldman 1986, 36). He goes on to protest against Quine’s infamous (1953b) attacks on analyticity, by insisting that there must be “some substance to the commonsense notions” of meaning and synonymy, that even philosophers who reject analyticity often perform something *like* conceptual analysis when they reason philosophically, and that presenting necessary and sufficient conditions is an indispensable approach to philosophical reasoning, even if it has a long record of failure (1986, 38–9).

In chapter 3, we will examine attempts to address Goldman’s first concern, and make naturalistic sense of analyticity. Suffice it to say that it is no small task. As to Goldman’s second concern, we will shortly discuss whether there is something sufficiently *like* conceptual analysis to do the philosopher’s task. This is particularly urgent, because Goldman’s third point about the indispensability of analysis in the face of its failure looks particularly implausible twenty years later, after the analytical debate over the Gettier problem has long fizzled out, and if any consensus has been reached, it is only that a new approach to epistemology is needed. Naturalists, now under the guise of "experimental philosophy," stress the diversity and cultural dependence of philosophical intuitions (Nichols, Stich and Weinberg 2003). Indeed it is arguable that the analytic "naturalists" whose roots are found in the Gettier problem are only accidentally related to naturalists like Quine, whose motivations were very different, as we shall find in chapter 2 and chapter 5.

If some version of analytic naturalism can be salvaged as a conceptual project, however, its doctrinal implication becomes apparent. Combining a successful analysis of "knowledge" (in terms of reliability, etc.) with results from cognitive psychology enables us to determine whether and to what extent human knowledge exists. Goldman thinks that his analysis at least permits us to accept that knowledge is logically possible, even if the analysis does not entail that such knowledge exists and doesn't permit a knock-down answer to skepticism (1986, 55–6). To know if we know would require that we know our beliefs to result from a reliable process, and it is logically possible to know *this* (56–7). Only our best psychology, not any analysis, can inform us as to whether that possibility is actual. It is at this point that objections of circularity usually enter, but Goldman has the option of noting that arguments for skepticism only arise because of conceptions of knowledge uninformed by reliabilism, conceptions that require ruling out Cartesian alternatives that are not relevant. Of course not all naturalists are as confident as Goldman about the power of cognitive psychology to deliver good news (Stich 1990). And if the success of this doctrinal project depends on the success of conceptual analysis, doubts about the latter could turn the former into a "degenerating research program."

Two-factor semantical naturalized epistemology

We need, therefore, to seek an approach to the conceptual project that is *like* traditional conceptual analysis, but not committed to the same substantive presuppositions about meaning and synonymy. Even if a philosopher is not tied to philosophic intuitions about the meaning of concepts like "knowledge" and "justification," it may profit him to *begin* with those intuitions as an entrée to a more sophisticated scientific theory, the results of which may or may not end up bearing much resemblance to the original intuitions. What counts for this kind of conceptual project is not so much allegiance to prior intuitions, but the predictive and explanatory power of the theorist's ultimate conceptualization. The current literature features proposals for theories of concepts supplanting the classical theory of concepts drawn on by conceptual analysis, and these proposals are relied upon, implicitly or explicitly, for alternative formulations of naturalized epistemology. I will mention two such theories of concepts, and some paradigm applications in epistemology.

The classical theory of concepts drawn on by conceptual analysis held that concepts expressed conjunctions of necessary and sufficient conditions, which could be discovered by the introspective reflection of the theorist. This theory was called into question by the Twin Earth thought experiments, which seemed to indicate that meaning of concepts could not be "in the head," because the reference of a term like "water" seems to vary in relation to the environment in which it is originally deployed (whether it is an environment containing H_20 or XYZ). A recent view of concepts seeks to capture the insight of these thought experiments, while also preserving an element of the classical view. These "two factor" or "causal-descriptive" theories urge that one factor of meaning is determined by *a priori* factors, while a second is determined by external aspects of the natural or social environment. In the view of Frank Jackson (1998), for example, we begin with a description of water as a clear, liquid stuff found in rivers and streams around here. We *need* to grasp at least this much, if ever we are to eventually *discover* the reference of "water" in the external world (either H_20 or XYZ). Importantly, we may end up revising our concept of "water," but we need to appeal to our intuitions about it before we can ever make that discovery. Other two-factor theories are even more unabashedly naturalistic than Jackson's, and urge that the descriptive component of reference is not *a priori*, but a manifestation of background theoretical knowledge acquired through ordinary empirical means (Boyd 1991; Stanford and Kitcher 2000)

How might the two-factor view of concepts be implemented in naturalized epistemology? One theorist who seems to be implicitly committed to the view is Philip Kitcher. In his essay "The Naturalists Return" (1992), Kitcher considers Goldman's reliabilism to be a holdover of analytic epistemology, and claims that while "reliabilism gives a promising *start* to formulating a meliorative naturalistic enterprise," it is "not the panacea for the problem of analyzing justification" (69). He believes that when analytical naturalists define ideal standards of justification in advance of inquiry, they invite skepticism and fail to shoulder the proper task of epistemology. Goldman's reliabilism, treated as an analysis of knowledge, invites counterexamples of true beliefs caused by reliable processes in a bizarre manner, for instance. It is always possible to refine definitions to better capture our intuitions about knowledge, but this does little to improve our understanding of worthwhile cognitive goals or improve our ability to reach them. What Kitcher means by the "meliorative project" is precisely the kind we might guess to be recommended by a two-factor approach to reference[1]:

> Traditional epistemology has an important meliorative dimension. Bacon and Descartes were moved to epistemological theorizing by their sense of the need to fathom the ways in which human minds can attain their epistemic ends. If analysis of current concepts of rationality and justification, or delineation of accepted inferential practices, is valuable, it is because a clearer view of what we now accept might enable us to do better. Conceptual clarification has a role to play in advance of inquiry, even when we understand that our current concepts might give way to improved ones. (64)

Kitcher speaks here of the meliorative project of traditional epistemology, but it is clear from the rest of the essay that he sees naturalism as sharing the tasks of traditional epistemology, if not the means.

How, on Kitcher's view, do we come to understand these worthwhile cognitive goals and assess our prospects of achieving them? He would implement the doctrinal project of naturalized epistemology by looking to the history of science, and more fundamentally, to our evolutionary heritage. In the course of examining our actual cognitive practices, and the basic equipment we inherited to undertake them, we may discover that achieving our cognitive goals is not always consistent with our *a priori* epistemic standards. We may find that we need to replace rather than

[1] Stanford and Kitcher (2000) develop an explicit two-factor theory of reference.

analyze the dichotomies of "rational/irrational" or "justified/unjustified," out of the need to give a richer portrait of factors contributing to the limited human animal's achievement of its cognitive goals.

Kitcher is aware, of course, that not all naturalists would find epistemological solace in an examination of the history of science or in the human evolutionary heritage. The bulk of his doctrinal studies concentrate on answering their worries. These skeptics might doubt, for example, that the cognitive equipment of our ancestors needed to be geared towards the acquisition of significant truths in order for the race to evolve successfully. But even if our ancestors developed some remedy to possible evolutionary shortcomings, the more serious naturalist challenge to the possibility of outlining the means and ends of human cognitive progress is that posed by Quinean and Kuhnian underdetermination arguments. These suggest that science has not developed by a series of logically-sanctioned steps aimed at an ultimate cognitive goal, but instead by a series of paradigm shifts that could have been otherwise, because of pragmatic decisions about auxiliary hypotheses, etc. Kitcher believes that the only response to this challenge is to examine the historical record even more carefully, to show that instances of underdetermination are not as pervasive as critics suggest. (In the final chapter of this dissertation, we will return to the topic of the underdetermination, which underpins some of the most basic naturalistic assumptions—a point Kitcher does not seem to fully appreciate.) Kitcher also believes he can examine the history of science to answer persistent objections from Larry Laudan (1984) and to show that the putative diversity of historical scientists' goals can be reduced to "a single, compelling, conception of cognitive value," which Kitcher calls "significant truth" (1992, 102). Kitcher delivers a ground-level examination of these very questions in his exhaustive treatment, *The Advancement of Science* (1993).[2]

Another naturalized epistemologist, Hilary Kornblith (2002), subscribes to the same conceptual project as Kitcher, but goes further still. Appealing explicitly to Boyd's two-factor semantics, Kornblith argues that the epistemologist's reference to knowledge can be understood as

[2] For more on the debate between Laudan and Kitcher, see Rosenberg (1996).

reference to a *natural kind*, understood on Boyd's model of natural kinds as causal homeostatic mechanisms. A homeostatic mechanism is a natural cluster of highly correlated properties or elements, the combination of which promotes a self-reinforcing stability, such that predicates describing the cluster are readily projectible. The combination of hydrogen and oxygen atoms in the water molecule is a good example.

Kornblith thinks that knowledge is a natural kind like water because cases of knowledge "have a good deal of theoretical unity to them" rather than being a "gerrymandered kind" constructed by human convention (10). The "theoretical unity" of knowledge is first understood by reference to the theoretical unity of *belief*. Kornblith looks to animal ethology's extensive use of intentional idioms to describe, explain and predict a variety of animal behavior. Even ants returning to the colony seem to "represent" their direction and distance traveled. More sophisticated animals exhibit genuine beliefs and desires, when the information represented comes to form a stable product available for multiple uses, depending on the animal's desire. Kornblith understands knowledge as a species of belief, and adds that it features an extra dimension of explanatory/predictive value, also recognized by animal ethologists. Whereas the actions of individual animals could always be easily explained by reference to mere beliefs, explaining how it is their species possesses the cognitive capacities that permit successful interaction with the environment requires the appeal to *reliable* belief-forming processes, i.e. knowledge. In short, nature has selected these cognitive capacities for their survival value, which in turn ensures the perpetuation of the capacities themselves (57–9).

Kornblith's natural kinds-oriented conceptual project has important doctrinal implications. To show that organisms really do possess the relevant reliable capacities, he must answer critics like Brandom (1998) who allege that judgments about reliability vary in relation to the scope of the organism's environment, and theorists may circumscribe environments arbitrarily, according to their interests. Kornblith (2002, 65–9) responds that the concept of an environment is itself a technical concept of ecology, one that is just as naturalistically respectable as many used by biologists.

Knowledge, then, is specifically an *ecological* natural kind. Kornblith must also oppose popular positions in epistemology according to which animals cannot possess knowledge or beliefs, because both concern the essentially social practices of giving and asking for reasons (69–102), and because knowledge requires a kind of self-conscious reflection of which animals are incapable (89–136).

As we have progressed from analytic naturalism to Kornblith's two-factor natural kinds naturalism, we have become less focused on the *concept* of knowledge and more focused on the *metaphysics* of knowledge itself. He goes the furthest here, seriously downplaying the need to appeal to philosophic intuition. Responding to Goldman's (1993) contention that naturalized epistemology should at least describe our epistemic "folkways" (our inherited intuitions about knowledge) before engaging in object-level study, Kornblith notes that in chemistry, we do not bother cataloguing folk chemistry; instead we "can simply skip straight to the project of understanding the real chemical kinds as they exist in nature." He concludes that "we should take seriously the possibility that a similar strategy might be equally fruitful in epistemology" (2002, 19). Arguably the next version of naturalism would seem to push Kornblith's suggestion to the extreme, avoiding discussion of concepts entirely and going straight to the metaphysics of knowledge.

Epistemic supervenience naturalized epistemology

In an influential critique of Quine's naturalized epistemology—and cognizant of Quine's antipathy towards conceptual analysis—Jaegwon Kim (1988) proposes a method of formulating epistemological criteria that avoids controversial reliance on philosophical accounts of meaning. Utilizing a concept he has developed in detail largely in connection with topics in the philosophy of mind, Kim argues that it must be that epistemic properties *supervene* on natural ones:

> [I]f a belief is justified, that must be so *because* it has certain factual, nonepistemic properties, such as perhaps that it is "indubitable", that it is seen to be entailed by another belief that is independently justified, that it is appropriately caused by perceptual experience, or whatever. That it is a justified belief cannot be a brute fundamental fact unrelated to the kind of belief it is. There must be a *reason* for it, and

> this reason must be grounded in the factual descriptive properties of that particular belief. (399)

A number of other philosophers, including Van Cleve (1985) and Sosa (1980), have endorsed the notion of epistemic supervenience, without necessarily seeing it as a naturalization proposal. Although Kim is widely recognized as a critic of Quine's naturalism, his critique acknowledges the viability of naturalistic projects rivaling Quine's, such as Kitcher's and Goldman's (Kim 1988, 394–9). His own supervenience proposal, in fact, can be transformed into a kind of naturalism, provided that the properties that epistemic properties supervene upon are themselves natural properties, and also provided that the nature of the supervenience relation itself is naturalistically respectable.

Speaking loosely, supervenience is the determination of a higher level property by a lower level property. To say that higher-level property A supervenes on lower-level property B is to say that any two objects which do not differ in lower-level B properties must not differ in their higher-level A properties. Or: there cannot be a difference in A properties without a difference in B properties. Supervenient A properties must have *some* subvenient B properties of some type or other, but if anything has these subvenient B properties, the supervenient A properties must obtain. The nature of that "must" is of some importance. The strong notion of supervenience needed to support a *determination* relation between B and A properties requires some kind of necessity. At one point in his discussion of supervenience of the mental, Kim's favored option is to find a kind of *nomological* necessity (1985). If there is a lawlike relationship between B and A properties, that would secure the necessary strong supervenience. We will discuss this view of necessity in chapter 2.

With the concept of supervenience in hand, Kim seems to have formulated a metaphysical stand-in for the conceptual project in epistemology, and can proceed to look for answers in the doctrinal project. He can search the relevant science to see if a lawlike relationship does exist between any properties and epistemic properties. In an essay on the supervenience of the *mental* on the physical, he considers the possibility of psycho-physical laws in the context of the problem of the multiple realizability of the mental. He proposes that the physical instantiation of these psychophysical laws may consist of lengthy disjuncts of distinct properties. Whether science could ever uncover or deal with laws of this type is not clear. To the extent that epistemic properties are themselves dependent on doxastic ones, the same problem may apply to epistemic supervenience.

The nomological supervenience concept is at best a placeholder for scientific discoveries waiting to be made. To the extent that it requires the discovery of nomological relationships, it may draw strength from discovery of the very kind of homeostatic mechanisms that Kornblith believes animal ethology to have catalogued. Indeed if supervenience requires a notion of nomological necessity, there may be little difference between Kim's and Kornblith's views in the end. Later (2005), Kim appears to rely on a conceptual form of necessity. Either way, supervenience has affinity to conceptual projects we have already considered.

Common to Goldman, Kitcher, Kornblith and Kim is the conviction that knowledge *really is something*. Consequently they look to the natural sciences to "uncover" knowledge of what that something really is. But this is not the only possible naturalistic approach to answering skepticism. It is possible to affirm the truth of statements concerning knowledge without being ontologically committed to the substantive *existence* of knowledge-stuff. This possibility is one that has been explored by the next category of naturalized epistemology, one that has not always been recognized as such: deflationary naturalism. In discussing this next category, however, we enter into the realm of what I call "pessimistic" naturalized epistemology.

Pessimistic naturalized epistemology

Dividing philosophical views according to the labels of "optimistic" and "pessimistic" is, of course, loaded with value judgments. An optimistic expects success; a pessimist, failure. The present category of pessimistic naturalized epistemologies counts as pessimistic only insofar as they expect failure to achieve traditional epistemological goals. But these epistemologies are not absolutely pessimistic: they believe that their proposals offer alternative goals that can be readily achieved. I can only state here that I myself happen to side with (most of) the goals of traditional epistemology, and for this reason I am exercising the privilege of categorizing epistemologies relative to that position. At the end of the dissertation, I hope to have established that traditional epistemological goals—including some of the traditional means to these goals—should not be abandoned for the reasons naturalists are wont to abandon them. So hopefully the present categorization will prove to be useful.

Deflationary naturalized epistemology

Deflationary views in philosophy are generally concerned with explaining how one might affirm a type of philosophic truth without being committed to the existence of substantive properties related to predicates expressed in those truths. The classic example is the deflationary view of truth, which holds that the meaning of the truth predicate is exhausted by the disquotational formula: "Snow is white" is true if and only if *Snow is white*. This conception avoids the commitment to a substantive truth relation, and consequently avoids thorny metaphysical questions about the nature of correspondence or of facts to which truths must correspond. For some time now, deflationary views of "knowledge" have also been available, particularly from the contextualist wing of epistemology.[3] Until recently, however, it has not been obvious how deflationism might also count as a form of *naturalism*.[4] If knowledge is not a substantive existent, what about it would scientists have to study?

One clue is offered by Huw Price (2004). Speaking of philosophic issues apart from epistemology, Price notes that we can make a distinction between *object*-naturalist and *subject*-naturalist approaches to central concepts in these fields. The object-naturalist is concerned with discovering the substantive properties to which philosophic concepts refer, and as such employs the methods of natural science to discover them. The naturalized epistemologies we have considered so far surely count as object-naturalist. But the subject-naturalist is not so much concerned with substantive

[3] See Pritchard (2004) for a survey of prominent deflationists about knowledge, including Sartwell, Foley, and Williams.
[4] Henderson (1994) actually considers contextualism as a kind of naturalism, but not for the reasons I outline.

properties as he is with subjects' *use* of philosophic *terms*. The subject-naturalist in epistemology, then, would be primarily concerned with human use of the term "knowledge." This, as it happens, is the celebrated project of the contextualists.

An excellent case in point is Michael Williams (1996). Williams himself characterizes his position as deflationary (111–3), drawing explicit inspiration from Quine's deflationism about truth. For reasons we will discuss later in chapter 5, Williams later critiques Quine's views on naturalized epistemology (254–65). Nevertheless, his own deflationary view may count as a form of subject-naturalism, if Price's conception here is useful. He is surely no object-naturalist: contrary to Kornblith, he denies that knowledge is anything like a "natural kind," or any *thing* at all, denying the position he calls "epistemological realism." Williams's motivation for adopting this position emerges out of his critique of traditional epistemology. He has argued that skepticism is a consequence of foundationalism, in particular the view that our beliefs have foundations in the senses, and a consequence of the "totality condition," the idea that all of our knowledge can be assessed at once. When the skeptic considers these possibilities, he loses confidence in the possibility of sensory foundations, and in doing so loses confidence in the totality of knowledge. Williams urges that we abandon foundationalism and the totality condition in order to avoid the problem of skepticism. But this solution to skepticism is very dissatisfying: without foundationalism, we crave some other assessment of the source of our knowledge. Williams, therefore, takes it upon himself to explain what is wrong with the craving in the first place. His main response is that knowledge is not an object of theory in need of any explanation.

Williams thinks that it may be true that we know many things in the proper contexts, but that this is not in virtue of anything in common among the cases called "knowledge." To show that there is something significant in common among such cases, one would need to demonstrate that cases of knowledge have a kind of "theoretical integrity" (103). But our beliefs—to say nothing of our knowledge—are not *topically* integrated. We do not store them in the form of a single, allencompassing axiomatized system. All that remains is the possibility that they are *epistemologically* integrated, i.e., subject to the same constraints, tracing from the same sources. Of course Williams believes that foundationalism only leads to skepticism, so the claim that knowledge exhibits this kind of integrity is in danger of giving "knowledge" the status of a theoretical term (like "phlogiston") that fails to refer if the theory behind it is false. There are also terms such as "table" or "heat" whose reference is thought to be fixed pre-theoretically or theory-independently. But Williams can find no reason to think "knowledge" functions in the same way (109–10). In chapter 4, after considering evidence about our formation of the concept of "know" that casts doubt on deflationism about "belief," we will return to the question of the pre-theoretical integrity of "knowledge."

For this reason, Williams thinks all we can hope for from epistemology is a deflationary account of knowledge:

> A deflationary account of "know" may show how the word is embedded in a teachable and useful linguistic practice, without supposing that "being known to be true" denotes a property that groups propositions into a theoretically significant kind. We can have an account of the use and utility of "know" without supposing that there is such a thing as human knowledge. (113)

This is as close as Williams comes to stating a conceptual project for his epistemology. Unlike previous object-naturalists, he is not concerned with the question of what knowledge *really is.* He is primarily interested in the concept itself, and even then, mainly the *word.* The naturalistic investigator can then make use of this proposal for the conceptual project, by examining our actual linguistic practices to see if they stand up to Williams's contention that our attributions of knowledge lack any obvious theoretical integrity. If the investigator determines that this is true, this amounts to Williams's version of the doctrinal project in epistemology: by debunking knowledge as a natural kind, he will have dissolved our craving for epistemological explanations, and in doing so he will have shown why we can reject skepticism without needing to assess the totality of our knowledge.

We need, then, to briefly describe the kinds of investigations that would be relevant to supporting Williams's contentions about the linguistics of "knowledge." These, I think, would be little more than the familiar examples entertained by contextualists, concerning the shifting standards of justification from context to context. To defeat the foundationalist view of theoretical integrity, Williams believes that he need only show that there is never any single type of proposition which, in virtue of its contents, "will have an epistemic status it can call its own" (113). Here is a sample of the kind of ordinary survey that would support this:

> In both science and ordinary life, constraints on justification are many and various. Not merely that, they shift with context in ways that are probably impossible to reduce to rule. In part, they will have to do with the specific content of whatever claim is at issue. But they will also be decisively influence by the subject of inquiry to which the claim in question belongs (history, physics, ornithology, etc.)…. Not entertaining radical doubts about the age of the Earth or the reliability of documentary evidence is a precondition of doing history *at all*. There are many things that, as historians, we might be dubious about, but not these. (117)

To these "disciplinary" constraints, Williams also adds "dialectical" constraints and "situational" constraints, which derive from idiosyncrasies of conversational and evidential contexts. The role of context is even more important to Williams than simply providing evidence against foundationalist theory. It not only helps to show why we shouldn't worry about skepticism, but shows the consequences of what happens if we do. The disciplinary constraints he mentions not only keep us on task as historians and physicists, but *stop us from doing epistemology* (122). Paradoxically, it turns out that this "methodological necessity" is epistemically good for us, the epistemologist's questions about the totality of our knowledge actually cause us to *lose* our knowledge insofar as we share his doubts, insofar as they cause us to suspend the "interact relations with our environment that…are crucial to much ordinary knowledge" (358).

At one point, I almost decided to classify Williams's deflationism as a kind of "optimistic" naturalized epistemology. Williams does attempt to show how deflationism helps respond to skepticism by affirming our knowledge of many things. For all of this, however, his theory may still be deeply dissatisfying to the traditional epistemologist. He would blame this dissatisfaction on philosophers' lingering foundationalism, which he takes to be hopeless. (In the final chapter of this dissertation, we will revisit the question of whether better formulations of foundationalism might solve rather than cause the problem of skepticism.) In this respect he has much in common with the category we are about to examine, Quine's naturalized epistemology. As we shall see, however, the chief difference between Williams and Quine is that Quine is on the whole less concerned with dealing with the skeptic, and does not even make substantial theoretical *use* of the concept of "knowledge" in his proposal.

Quinean naturalized epistemology

Quine's position is the last we'll survey in this chapter, but it was also the first significant proposal for naturalized epistemology of the 20th century. Many of the previous views took inspiration from Quine, and take themselves to be following his research program. But even among those who revere his example, there seems to be a general consensus that Quine went too far, that his position represented an unreasonable abandonment of core elements of genuine epistemology. We can see why this is generally accepted by considering a widely-quoted representative passage from "Epistemology Naturalized" (1969a, 82-3) in which he describes the proper subject matter of naturalized epistemology:

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input—certain patterns of irradiation in assorted frequencies, for instance—and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence.

Quine's emphasis here on studying the actual psychological history of a subject's cognitive processes led many critics to regard him as abandoning the normative element of epistemology, the attempt to assess the *justification* of our beliefs. In this respect he seems very much like Williams. This has led scholars of the field to classify Quine's views as "eliminative" naturalism (Maffie 1990), or "replacement" naturalism (Almeder 1990; Feldman 2006). Many critics (e.g., Kim 1988) have seen this alleged abandonment of concern with normativity to be the chief flaw in Quine's position. Whether and in what sense Quine really does abandon normativity is, of course, a subject of some debate, of course, and in the next section and in the next chapter we will examine the question in some detail.

For the time being, however, I want merely to focus on the fundamental *methodological* uniqueness of Quine's proposal in relation to the other naturalisms so far surveyed. I have categorized naturalisms so far according to their particular conceptual and doctrinal projects, a distinction which comes from Quine himself. Quine, of course, would lend no quarter to *analytic* naturalism, as he is famously (or infamously) skeptical about notions of meaning and synonymy underlying the idea of conceptual analysis (Quine 1953b). Quine would not be any happier with the two-factor theory of concepts that underpins the second version of naturalized epistemology, on the grounds that it relies on a picture of reference inconsistent with his infamous inscrutability of reference thesis (Quine 1969c). Even the doctrine of natural kinds—drawing as it does on cherished examples from the natural sciences—feels the brunt of Quine's withering critique (Quine 1969b). Science based on natural kinds is "rotten to the core," says Quine—albeit a kind of rot necessary for progress to a better, more naturalistically respectable science based on mathematics (1969b, 133). Even the epistemologically minimalist doctrine of supervenience would irk Quine, owing to its essential reliance on the concept of "necessity." "Necessity" is an intensional concept, like "belief," that Quine judges to be incompatible with his naturalistic extensionalism (1953b). We will explore many of these Quinean objections to optimistic naturalism in chapters 2 and 3.

Without recourse to standard philosophical methodologies, by what means does Quine hope to naturalize epistemology? Here it is worth repeating that although the distinction between conceptual and doctrinal projects in epistemology is Quine's, it is not one that he applies to his own work. He does not begin by analyzing the concept "knowledge" and then use science to determine to what extent the concept is applicable. In fact Quine has little interest in the concept of "knowledge" to begin with.

He is more concerned to examine the relationship between "evidence" and "theory," where the latter is understood primarily *linguistically*, rather than cognitively. Surprisingly, he also has very little concern with assessing the *reliability* of this linguistic output. "Reliability" is a concept intimately connected to truth, but for Quine, epistemology is chiefly concerned with explaining the *pragmatic* value of our linguistic outputs, i.e. their facility in allowing the prediction of experience and subsequent control of our environment. At the same time, some of Quine's writings on epistemology appear to offer a naturalistic answer to skepticism (Quine 1974; 1975b). It is not immediately obvious how to reconcile this approach with Quine's pragmatism, but we will attempt to do it in chapter 5.

In the end, I will argue that Quine's naturalism is the most naturalistic of the naturalisms, and that if we want to evaluate naturalized epistemology in its most fundamental terms, we must evaluate Quine's most fundamental theses (this will be the subject of chapter 6). But before presenting my polemical strategy, I will survey one of the usual objections raised by prominent critics of naturalism, to show why I think my strategy raises concerns that are more fundamental.

A representative objection to naturalism: the normativity objection

Objections to naturalized epistemology vary with the variety of naturalism. Every version considered so far depends on a distinctive philosophic methodology—and on the philosophic presuppositions of that methodology—and any objection to this methodology would sensibly count as an objection to the associated epistemology. It is curious, however, how little most critics of naturalized epistemology focus on these basic philosophic considerations. Many instead focus on objectionable features and alleged internal inconsistencies of any approach that makes natural science the essential methodology of a discipline traditionally thought to be purely philosophical. In this section, I will survey one such objection that contains elements of many of the other typical objections, and so serves as a good representative. This is the objection that naturalized epistemology fails to do justice to a central aspect of serious epistemology: its status as a *normative* discipline. I will argue that traditionalist objections to naturalism raised on these grounds are not ultimately convincing. If the naturalists' basic methodological and philosophical presuppositions are left unaddressed, the normativity challenge can be answered through any number of artful dodges.

One of the most prominent critics to register the normativity objection to naturalism is Jaegwon Kim (1988). Kim specifically targets Quine's version of naturalized epistemology, but in a manner that could generalize to many of the other naturalisms we have examined (even though in a way his own position is naturalistic). Kim alleges that a naturalized epistemology such as Quine's aims only to describe or even explain how our beliefs are formed, using the resources of descriptive cognitive psychology. Kim assumes that Quine is only interested in a descriptive or explanatory effort, and *not* in the normative project of explaining how our beliefs are *justified*. There is something right about Kim's observation here: there is some conception of justification that Quine is not interested in discussing. As we shall see later in this section, however, this does not imply that Quine rejects the possibility of *any* normative role for epistemology. Whatever the proper interpretation of Quine, however, it is clear that this is a natural objection to the view that epistemology could somehow be based on science: since science is thought to be mainly descriptive and/or explanatory, how could science concern itself with a matter traditionally associated with philosophy, the question of right and wrong in the way of believing?

One critic of naturalism who has developed the normativity objection in more detail is Harvey Siegel. In one essay (1989), Siegel critiques the naturalized epistemology of Ronald Giere (1985; 1988). Giere's "evolutionary" naturalism attempts to "explain how creatures with our natural endowments manage to learn so much about the detailed structure of the world" (Giere 1985, 339–40). Siegel alleges that Giere's epistemology abandons concern with the *rationality* of science, and further alleges that the concept has "no place in a naturalized philosophy of science" (Siegel 1989, 366). Siegel (1989) objects that the scientific study of science, while unobjectionable in its own right, cannot answer traditional philosophical questions about science, questions such as "Is there a scientific method which warrants or justifies claims arrived at by its use?, Are there principles of theory choice which render such choice rational?, What are the epistemic relationships between theory and evidence, and between those two and truth?" (368). But Siegel contends that Giere has not shown these to be non-viable questions. So if naturalized epistemology is really incapable of examining questions of rationality, and these questions are viable, it is failing to deal with important, viable *philosophical* questions, and also failing to do the job of any philosophy of science worth its salt.

Of course Siegel's objection only works if we take it for granted that there is no room in the scientific study of science for questions of rationality. Giere's position on this question is qualified. He says that rationality "is not a concept that can appear in a naturalistic theory of science—*unless reduced to naturalistic terms*" (Giere 1985, 332, emphasis mine). As if to anticipate the suggestion that Giere or someone else could actually provide a naturalistic reduction of rationality, Siegel offers another objection also typically associated with traditionalist critics of naturalism. Science could never answer questions about rationality, he argues, because

> To answer [these questions] scientifically would be to beg the question—e.g., any answer to the question of the relationship between evidence and a justified theory, if arrived at scientifically, would depend upon exactly the same relationship between it and the evidence for it as it recommends for the relationship between any justified theory and the evidence for it. Because these general questions about the epistemology of science cannot be answered naturalistically without begging the question, they cannot be so pursued. (Siegel 1989, 369)

So there are two questions in need of answering. First, is it true that science has no resources to deliver an account of rationality? Second, is it true that any putatively scientific account of rationality would beg the question?

Regarding the first question about the ability of science to deliver an account of rationality, Giere evidently does intend to cash out his qualification that rationality might still be reduced in naturalistic terms. In his response to Siegel, Giere contends that there is a notion of rationality readily available to the naturalist: *instrumental* rationality:

> To be instrumentally rational is simply to employ means believed to be conducive to achieving desired goals. . . .Actions that do not fit the model are labeled "irrational". .

> . . [T]here is also a more *objective* sense of instrumental rationality which consists in employing means that are not only believed to be, but are *in fact* conducive to achieving desired goals.
>
> This latter, objective, sense of instrumental rationality provides the naturalistic theorist of science with ample means for making normative claims about science. These claims, however, are not autonomous but are grounded within science itself. It requires empirical research to determine whether a particular strategy is in fact likely to be effective in producing valuable scientific results. (Giere 1989, 380)

As an example of this kind of rationality, Giere mentions scientists' adoption of the continental drift hypothesis in the 1960s. Although the hypothesis had been widely regarded as implausible until that time, theorists reasoned that if true, the hypothesis would imply that strips of the ocean floor should exhibit distinctive patterns of magnetism. Upon finding that these patterns existed, scientists choosing to adopt the continental drift hypothesis would be making an instrumentally rational decision: treating successful novel predictions of hypotheses as evidence for the same hypotheses has, historically, been an effective means of adopting models for real processes.

Before we address the second question, about whether Giere's proposal would beg the question, it is useful to examine Siegel's likely response to the naturalist's answer to the first question. In a later essay, Siegel (1990) addresses Laudan's similar (1984, 24, 34) proposal for how normativity can be naturalized via instrumental rationality, by considering an example of Laudan's about how scientists might discover the preferability of double-blind to single-blind experiments in medical research. Siegel contends that evidence of the history of the different types of experiment would establish the preferability of double-blind experiments *only* if investigators value learning the genuine effectiveness of drug treatments, as opposed to, say, learning the *approximate* effectiveness of drugs in the quickest and cheapest manner possible. If investigators valued approximate effectiveness more, they might well side with the value of single-blind experiments. Siegel's point is that preferring double-blind methodology is the only acceptable preference, and that this preference is justified "not instrumentally, but *epistemically*: double blind experimentation provides better *evidence* for a drug's efficacy than single-blind experimentation, because it controls for an additional source of possible error" (1994, 301). In suggesting this difference between epistemic and instrumental rationality, Siegel is relying on some of the same ideas behind his charge that science begs the question in attempting to provide a scientific account of rationality: that charge presupposes the possibility of a non-scientific, presumably *a priori* standard for assessing the justificatory value of evidence—a presupposition that we will examine shortly. Another way of looking at Siegel's objection is that even if we do conceive of the justification of double-blind experimentation instrumentally, we can do so only if we have offer *a priori* justification of the goal of truth, rather than the goal of approximate truth found in a cost-effective way. His assumption is that naturalists can naturalize instrumental rationality *only*, not the intrinsic or categorical rationality of the ends of inquiry. Should we accept this assumption?

Siegel realizes that naturalists such as Laudan (1987, 29) have offered proposals for naturalizing not only epistemic means, but epistemic ends as well. Laudan's proposal claims, for example, that certain ends can be disqualified by science if they are "utopian" and unachievable by human beings, or if their truth conditions cannot be specified, or if the means to their achievements cannot be specified. But Siegel insists that the unachievability of a goal is only a reason to disqualify it if we presuppose an instrumentalist conception of rationality (which would beg the question) (Siegel 1990, 307). Why he thinks this is unclear: it seems that Laudan need only rely on the conventional wisdom that "ought" implies "can" (see also Kitcher 1992, 83–7). To say that X is an instrumental value is to say that that X is valuable only if it achieves value Y. This is not the same as saying that X is valuable only if it *can be achieved*, full stop. A better objection to Laudan's position, I think, would be to say that just because science can help *disqualify* ends in this manner does not imply that it can help *establish* them in the first place. The challenge for the naturalist is to show how this might work, and perhaps Siegel is presupposing that the ends of inquiry can be justified only by appeal to *a priori* considerations. We shall examine that presupposition presently.

Laudan (1990a, 316) observes that a central assumption of Siegel's objection is that "there is something called epistemic rationality or epistemic justification which stands *outside* the analysis of ends/means connections." As we have seen, this presupposition is relevant not only to Siegel's last objection to the possibility of naturalizing ultimate ends, but also to his more general contention that any such attempt would beg the question. In his article, Laudan never purports to offer a naturalization of ultimate ends. He contends that all rationality is instrumental in relation to desired ends, *whatever those ends happen to be*. This is certainly a position naturalists could accept provisionally in the absence of establishing ultimate ends of their own: they could simply conclude with Hume that reason judges not of ultimate ends, that it is merely the slave of the passions. The more important point that Laudan makes, however, is that even though Siegel's objection relies on the existence, or at least the possible existence, of some *a priori* standard of rationality, he never tells us what it is or how we can discover it. Laudan mentions the example of the thesis that successful, novel predictions tend to confirm their associated hypotheses. Philosophers have sought an *a priori* standard according to which this counts as evidence, but have failed. All that philosophers managed to determine was that "we could attempt to ascertain whether theories which had proved themselves successful over the long run were theories that could have successfully satisfied the demands of this rule" (321). Indeed, naturalists such as Quine (1953b) would insist that there is no way in principle of defining objective standards of evidential confirmation, because theory is underdetermined by evidence, and theory choice is driven, ultimately, by pragmatic factors. Laudan himself would object to Quine's underdetermination thesis (in our final chapters, we will see why), but the present point is that even attempts to define standards of evidence by reference to the history of science (like Laudan's) face an uphill battle against the underdetermination thesis—to say nothing of attempts to define these standards *a priori*. As far as I can tell, Siegel's subsequent (1996) response to Laudan does nothing to address this problem. No where does he attempt to tell us what the *a priori* standard of evidence is or how we are to discover it. As far as we are concerned now, therefore, Laudan's view that all rationality is instrumental may be true: Siegel has offered no reason to think that reason is anything other than a slave to the scientist's passions.

Of course we should not discount entirely the possibility that other naturalists might still succeed at naturalizing epistemic *ends* themselves. Laudan does not consider this possibility, but other naturalists have. We have already discussed how, in response to Laudan's (1984) contention that the history of science is scattered with a plethora of different ends of inquiry, Kitcher (1993) argues that, with appropriate care, these can be reduced to a single end, the pursuit of "significant truth." Kitcher argues that current scientific theories survive because of having undergone a process of "natural" selection, in which theories with the greatest predictive and explanatory power are the ones that "survive." If Darwinian natural selection can explain the teleological function of biological processes in terms of adaptive success (Wright 1976), it seems reasonable that Kitcher could offer a similar quasi-Darwinian explanation of the teleological function of scientific theories, and underwrite the "significance" of his significant truth (Rosenberg 1996, 18). Of course whether or not this explanation is consistent with *realism* about the products of science is a matter of some controversy; it depends in large part on the verisimilitude of the stock of hypotheses scientists begin with before they eliminate all but the "fittest." Whether science can account for the original verisimilitude of our ancestors' theories as a product of cognitive evolution is a matter of some controversy, by Kitcher's own admission (1992, 92–3; Rosenberg 1996, 23).

Whatever the nature of the ends of science, a naturalist's formulation of it need not always cohere with our pre-theoretic conception of rationality. As we discussed in some detail while examining proposals for "analytic" naturalism, a naturalist may be content to begin with intuitions about rationality, without remaining faithful to them at the end of investigation. The folk intuition of rationality may not have the same predictive and explanatory power as the conception fully informed by the history of science and cognitive psychology.

In any case, as Alexander Rosenberg (1996, 25) observes, naturalists can characterize the end of science in terms even more generic than "significant truth": without resolving the realism/anti-realism debate, they can point to the ultimate end of science as nothing more than "prediction and control" (of experiences, if not of real entities). Why then are prediction and control to be taken as ultimate ends? Rosenberg considers the possibility that these might also be justified through their adaptive value, via Darwinian natural selection. But he thinks that this would go too far and lead to a vicious circularity. His alternative runs as follows:

> The only way naturalism can avoid [conceding the possibility of incommensurable goals of inquiry] is to show how those who reject naturalism in fact willy-nilly embrace prediction and control as the ultimate epistemic value, despite their claims to the contrary. . . . It must claim that its competitors' rejection of prediction and control as the ultimate aims of enquiry is belied by their own actions, choices, decisions, and provisions. (26–7)

But this self-refutation strategy seems quite dubious. Perhaps non-naturalists in everyday life can *appear* to embrace predictive and explanatory value without its being the standard of their inquiry. And it seems likely that non-naturalists could always formulate some alternative characterization of their predictive/explanatory behavior, as most theorists charged with self-refutation are usually able to do. Naturalists respond to non-naturalists like Siegel in the same way. If Siegel claims that naturalists have to rely on non-naturalistic conceptions of evidence, naturalists can simply provide a naturalistic explication of their practice.

The more important question to ask is: is Rosenberg correct that explaining the value of prediction and control via Darwinian theory would be viciously circular? Certainly it would be circular in some sense: the question is whether the circularity would be *vicious*. For as much as Rosenberg appreciates the willingness of naturalists like Giere, Laudan and Kitcher to embrace science and disavow the *a priori* approach to epistemology, he seems to ignore Quine's reason for seeking a naturalization of epistemology in the first place: the fact that attempts at ground-up reconstructions of science have always seemed to fail, and that foundationalism as a doctrinal project has long been a dead-end.[5] Yet like Siegel and other non-naturalists who charge naturalists with begging the question, Rosenberg seems to presuppose that something like a rational reconstruction of knowledge ought to be possible, if not through the direct appeal to foundations, then at least through a coherentist-style

[5] And, Rosenberg (1999) has aligned himself explicitly with Quine's project.

transcendental argument from self-refutation. If we heed Quine's original intentions, however, we should demur at all requests to offer rational reconstruction. We should forget about creative reconstructions and simply settle for psychology. We should show how, assuming the best science of the day, we explain the origins of that same science. So if we can understand the value of prediction and control using Darwinian theory (a point that Quine himself endorses), then so be it. That is a substantial achievement, even if it is not in keeping with the goals of traditional epistemology, and merely an achievement of Quine's "pessimistic" naturalized epistemology. We should be pragmatic naturalists of this manner with Quine—that is, if Quine and all of the other naturalists are correct that foundationalism and all other forms of rational reconstruction are hopeless.

If I am correct in the above, naturalists have the resources to answer the normativity objection in a variety of ways. Non-naturalists who pursue this objection usually fail to appreciate the different methodologies furnished by the many varieties of naturalism I have surveyed in my earlier section. In essence, non-naturalists fail to realize that naturalists can offer a naturalization of rationality or normativity in the same manner that they seek to offer to naturalize knowledge itself. This reduction need not comport with our pre-theoretic intuitions about evidence or justification. It need only draw on the history of science or cognitive science in a way that serves a useful scientific purpose. Depending on which naturalist we side with, the naturalist's conception of normativity can bear approximate or only minimal resemblance to our original intuitions of the ends and standards of inquiry. It can do so and avoid charges of circularity, because these charges hold up only on the presupposition that *a priori* standards can be formulated, and the entire project of naturalism has been motivated, of course, by the conviction that they cannot be. If we want to refute naturalized epistemology, we will need to challenge that conviction. It is just such a challenge that I will pose in this dissertation.

Outline of the dissertation

In order to establish the thesis that we must challenge the fundamental philosophic convictions of the naturalists in order to undermine naturalized epistemology, I seek to identify those convictions by playing naturalisms against each other to see which of them will yield to the others under pressure. In the end I will argue that "optimistic" naturalisms must give way to "pessimistic" ones of the Quinean variety. I will establish this point in the first half of my dissertation.

In our next chapter, chapter 2, I will examine Jaegwon Kim's (1988) influential critique of Quine's naturalized epistemology. Kim is best known for having advanced the normativity objection to Quine's thesis, but his critique is actually more extensive than this. Kim also proposes that Quinean naturalism is not the only alternative epistemologists should consider, given the assumption that deductivist foundationalism is moribund. He proposes additional alternatives that we can now recognize as versions of optimistic naturalism, including reliabilism and his own supervenience view. In my arguments above, I alluded to the fact that these optimistic naturalisms may conflict with naturalism's methodological proscriptions against analyticity, modality, and other intensionsal notions as represented by Quine's basic theses. In my examination of Kim, I will argue in explicit detail that this conflict is real, and that pending the shouldering of a massive burden of proof, Kim and other optimistic naturalists must yield to Quine's basic position.

In chapter 3, I apply further pressure to optimistic naturalism, and argue that one of its core concepts, the concept of "belief" (of reliable "belief" formation fame) cannot itself be naturalized by the same rigorous standards that naturalists want to apply to "knowledge" and "justification," at least not in a way that yields a concept of "belief" usable by the naturalized epistemologist. I will demonstrate this by enumerating a variety of belief-naturalization proposals utilizing methodologies similar to the optimistic proposals enumerated above. I will argue that they all face the same methodological tensions as the naturalisms analyzed in the Kim chapter. In the end, whatever naturalization of "belief" the basic naturalistic constraints permit will longer describe the kinds of beliefs needed by epistemologists to account for the possibility of sophisticated scientific beliefs. In addition to cutting off this aspect of the optimistic naturalists' naturalization project, I believe that focusing on the concept of "belief" will also help us understand some of the more basic methodological shortcomings of naturalization proposals of any kind. In particular, I will return to the issue of naturalizing normativity, and show how the appeal to Darwinian theory can only yield limited satisfaction whenever applied to naturalizing philosophical concepts.

The difficulty in naturalizing a substantive concept of "belief" need not ruin the project of naturalizing epistemology, of course. In chapter 4, I will look at the attempt by pessimistic naturalists to *deflate* the concept of belief, by offering a subject-naturalization of the deployment of the *term* "belief" using a contemporary theory in philosophy of psychology, the simulation theory of Alvin Goldman and Robert Gordon. I will argue that the most naturalistic version of the theory, Gordon's, fails to account for all of the evidence, particularly in the domain of developmental psychology. In offering this critique, I will begin to move away from applying pressure to optimistic naturalism in favor of doing the same to pessimistic naturalism. Also of interest in this critique will be the fact that evidence from developmental psychology suggests that children understand *knowledge* before they understand *belief*. Understanding this will help us to understand how the concept of "knowledge" can have theoretical significance prior to the development of adult epistemological theory, a point that helps undermine Michael Williams's arguments against epistemological realism and case for epistemological deflationism.

I will begin the second half of my dissertation in chapter 5, by explaining in more detail the content of Quine's pessimistic proposal, first by showing that it is in fact pessimistic. There, I show how Quine's naturalized epistemology deals with traditional skeptical worries, not by trying to refute them in the ordinary sense, but by showing that traditional justificatory goals of epistemology must be abandoned and replaced by pragmatic ones. The resulting naturalized epistemology concerns itself with identifying the various steps (justified or otherwise) by which human subjects develop their current, instrumentally successful scientific theory. Furthermore, I argue that Quine's pragmatic approach to epistemology faces difficulties of its own, particularly because of challenges posed by more radical pragmatists who see no pragmatic basis for privileging natural science over other forms of human discourse.

Of course, the putative motivation for pursuing a pragmatic rather than a traditional route stems from the alleged failure of foundationalism and the inevitability of Quine's indeterminacy of translation, assuming the viability of the underdetermination thesis. In chapter 6, I argue that looking at the wider context of scientific practice—and at some specific scientific results—undermines the underdetermination thesis and draws attention to the possibility of a new foundationalism. I argue that the underdetermination thesis does not itself bear naturalistic scrutiny, in particular because of its reliance on a crude and unrealistic hypothetico-deductivist conception of confirmation. Having undermined underdetermination naturalistically, I show how to generalize this anti-skeptical strategy (itself inspired by statements from Quine about skepticism): whenever skeptics themselves assume points of science for the sake of *reductio ad absurdum*, anti-skeptics themselves have the right to make appeal to science to show how the *reductio* does not go through. I demonstrate that the classical Humean problem of induction can be partially resolved by appeal to a material theory of induction that recognizes diverse methods of confirmation practiced by scientists in different domains of fact. Finally, by appealing to psychological evidence regarding perception and concept-formation, I show how the regress of inductive justification can be terminated in perceptual foundations. At the same time, I argue that as skeptical problems become more dependent on questions about epistemological foundations, the problems become more and more philosophical and less purely scientific. This suggests that Quine is ultimately incorrect that skeptical doubts are prompted entirely by scientific problems. This means we cannot generalize his anti-skeptical strategy to solve all skeptical problems, but it also means that the naturalistic proposal to make philosophy continuous with natural science is not consistent with the fact that naturalism itself arises as a pragmatic solution to problems generated by non-naturalistic philosophic presuppositions.

The outcome of this anti-skeptical strategy robs naturalized epistemology of its *raison d'etre*. At the same time, the strategy also draws attention to flaws in traditional epistemology, including its inability to articulate a workable empirical foundationalist theory of justification. These flaws must be corrected if philosophers are to preserve the autonomy of their discipline.

CHAPTER 2

KIM'S CRITIQUE OF QUINE'S NATURALIZED EPISTEMOLOGY

Numerous critics of W.V. Quine's essay "Epistemology Naturalized" treat Quine's proposal to make epistemology a "chapter of psychology" as a proposal for abandoning normative epistemology (Quine 1969a). One of the most prominent critics making this contention is Jaegwon Kim. Kim objects that by merely describing the *causal* relationship between cognitive input and output, Quine's naturalism abandons the *normative* concept of "justification," the normative element of the concept of "knowledge", and therefore genuine epistemology (Kim 1988). Kim also urges that aside from the concept of "justification," even the concept of "belief" has a normative dimension, and that any epistemology wishing to dispense with normativity must also dispense with "belief"—a seemingly absurd consequence for naturalists who otherwise seem to be enamored of discussing reliable *belief*-forming processes.

There is, however, a serious question as to whether Quine's approach to normativity is a fundamental vulnerability in his position. In later writings, Quine denied that he intended to dispense with normativity (Quine 1986a; 1992).[6] Others have explained in detail the resources available to Quine for the purpose of naturalizing epistemic normativity (Sinclair 2004). Beyond the specific domain of Quine scholarship, proposals for naturalizing normativity in terms of special facts about biological or sociological causation abound (Giere 1988; Laudan 1984; Laudan 1990b; Rosenberg 1996; Kornblith 2002). These naturalization proposals are controversial, of course, but their existence demonstrates that one cannot challenge naturalized epistemology simply by assuming an opposition between the causal and the normative.

Surely Kim is aware of the possibility that normativity might be naturalized, and although his critique is not often remembered as such, it involves more than just the normativity objection. Kim

[6] Recent defenders of Quine have also offered explanations for why critics erroneously came to read Quine as rejecting normativity (Johnsen 2005).

also explores what he takes to be Quine's motivations for naturalizing epistemology in the first place, and argues that they are inadequate. In particular, Kim argues that the failure of traditional deductivist foundationalist epistemology does not, as Quine believes, leave Quinean naturalism as the only alternative. To support this, he points out non-Quinean alternatives to traditional epistemology, a variety of mainly contemporary theories that propose new, non-deductivist conceptions of justification. Insofar as these any of these alternatives is viable, it seems that Quine may have overlooked the third way between tradition and naturalism. The question, then, is whether the alternatives *are* viable.

I would suggest that this question about the viability of non-Quinean alternatives, rather than the normativity issue, is the more fundamental question for naturalism: naturalists have too many tools at their disposals to solve or dissolve the normativity problem. This chapter, therefore, will examine the viability of Kim's alternatives. As it happens, there are distinctively Quinean objections to be registered against each of them. But Kim does not anticipate most of these objections or the power of the distinctively Quinean principles that empower them, principles such as the underdetermination thesis, the indeterminacy thesis, and extensionalism.

To demonstrate this, I will begin with a brief review of Quine's arguments in "Epistemology Naturalized." I will then summarize the non-Quinean alternatives to traditional foundationalist epistemology Kim considers, and show how Quine might refute each by reference to his basic principles and methodology. In particular, I will discuss the methodological concept Kim relies on to propose many of his alternatives, the concept of supervenience, and show the same Quinean objections might be applied to it. I will end by examining one likely element of a would-be supervenience base for epistemic properties—belief states—and raise similar problems.

In pointing out Quinean responses to Kim's alternatives, I do not intend to defend Quine to the bitter end or reject Kim. I myself am skeptical about naturalized epistemology, and sympathize with the spirit of Kim's objections. But I think that a fair criticism of Quine must be rendered at the most fundamental level. If deep Quinean principles make it difficult to propose alternatives to his epistemology, the only way to challenge his epistemology is to challenge those principles. Understanding this will help to illustrate the deep difference between Quine and his rivals.

“Epistemology naturalized” in brief

Kim focuses his criticism largely on Quine’s essay “Epistemology Naturalized.” It is important to note, however, that this essay was merely a summary of a project Quine had already been pursuing for years. As early as “Two Dogmas of Empiricism,” Quine had sought to undermine the positivist insistence on an analytic/synthetic distinction that had segregated philosophy from science (Quine 1953b). He rejected the distinction, not just because philosophers had not drawn it clearly, but on principled grounds deriving from confirmation holism. According to this view, only blocks of theory as a whole can be confirmed or refuted, so any individual statement can be held true or revised come what may; therefore, analytic statements are not a distinctive type of statement that is confirmed no matter what. With this rejection of analyticity, he held statements of science and statements of mathematics or logic to be on equal epistemological footing. This would presumably also include statements of philosophy—and thereby epistemology. In *Word and Object* (1960, 3), Quine told us that “the philosopher and the scientist are in the same boat,” and it is it is arguable that *Word and Object* is his major treatise on the subject of naturalized epistemology.

Fast-forward, then, to “Epistemology Naturalized.” Quine begins his essay by noting a parallel between mathematics and epistemology.[7] He notes that traditionally, both mathematicians and epistemologists pursued so-called “doctrinal” and “conceptual” projects. Doctrinally, epistemologists had hoped to identify ultimate justifiers of empirical knowledge, usually somewhere in the data of the senses, just as mathematicians had sought to prove mathematical truths by reference to self-evident first principles. Conceptually, epistemologists had hoped to clarify the meaning of the terms of

[7] It is actually more than a simple parallel, as Quine believes epistemology to include the study of the foundations of mathematics.

empirical knowledge in sensory language (particularly the notion of "body"), just as mathematicians had sought to define mathematical concepts in logical and set-theoretical terms.

In mathematics, the conceptual project aimed at assisting the doctrinal: defining obscure terms by reference to clearer ones could help to establish the relationship between obscure mathematical claims and more obviously true ones. Since 20th century mathematicians discovered that mathematical concepts could not be reduced to exclusively logical ones, but to logic and more obscure set theoretical terms, the mathematical doctrinal project stalled. In epistemology, empiricists like Hume sought similar cooperation between the conceptual and the doctrinal projects. By conceptually identifying bodies with types of sense impressions, Hume enjoyed limited doctrinal success in grounding some singular statements about bodies. But epistemologists made progress on the conceptual project only to the extent that, like the mathematicians, they resorted to the use of set theory (to expand their resources from simple impressions, to sets of impressions, etc.) and contextual definition (defining terms by translating whole sentences containing the term). The empiricist doctrinal project stalled because of the failure to ground generalizations and singular future tense statements.

Quine's proposal for naturalizing epistemology grows out of his contention that even Carnap's modest project was inadequate. According to Quine, even Carnap's work did nothing to advance the doctrinal "quest for certainty" in the face of Humean problems. The most it could achieve, doctrinally, was a clarification of the nature of sensory evidence through a rendering of the conceptual meaning of scientific discourse. Given this, Quine wonders what would lead Carnap to pick one of many possible "creative reconstructions" of scientific discourse over a psychological description of the actual link between evidence and theory. At best, Quine imagines that reconstruction could at least *legitimize* scientific discourse in the sense of showing that its terms could be eliminated in favor of the putatively respectable concepts of logic and set theory, if it could be translated in these terms. This would be a kind of exculpatory legitimization, if not a justificatory one.

But famously, Quine thinks this kind of translation is impossible. He notes that Carnap himself admitted this when he later proposed specifying "reduction forms" which merely gave implications of the sentences in question, rather than translational equivalences. Abandoning translational equivalence—which at least enabled the kind of exculpation I've described above—left Carnap's rational reconstruction without any advantage over psychology. And, Quine thinks that the failure of translation, had to happen if basic principles of his philosophy were true. Revisiting themes from "Two Dogmas," he invokes Duhem's confirmation holism (the idea that only theories, not individual statements are ever confirmed or falsified) along with Peirce's verificationism (the equation of meaning with method of empirical confirmation) to argue for meaning holism (the assertion that only blocks of theory, not individual statements, have meaning). Revisiting *Word and Object,* he quickly derives the indeterminacy of translation from meaning holism: if individual statements have no meanings to call their own, there are no facts of the matter to determine the correctness of translation of individual statements. Hence, the failure of Carnap's efforts at translation, and the abandonment of the last advantage of reconstruction over psychology. So the failure of the empiricist conceptual project implies the complete abandonment of even a modest (exculpatory) doctrinal project.

It is only after this development that Quine (1969a, 82–3). comes to his famous conclusion about the need for a naturalized epistemology:

> [E]pistemology still goes on, though in a new setting and a clarified status. Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input—certain patterns of irradiation in asserted frequencies, for instance—and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence.

Quine's reference to a "new setting" and "clarified status" is remarkably precise. In what follows in the essay, he specifies what he means by each.

The "new setting" of epistemology is psychology. Quine thinks epistemology is "contained" in natural science, given that the subject of epistemological study is a physical, human one. This implies, for example, that when deciding which factor to count as an observation, as epistemologically prior, we should choose whatever is closest, causally, to sensory receptors, not whatever is related to awareness of sensory atoms or *Gestalten*. Defining observation by reference to awareness or sensory atoms would be needed only for a justificatory or exculpatory approach. Observation sentences, then, are simply those about *bodies,* because these are the statements most automatically and universally cued to identical sensory stimuli.

Understanding this "new setting" of epistemology helps identify its "clarified status." First, since we are not occupied by the foundationalist concern of justifying science in terms of awareness of anything, there is nothing circular about the psychological setting of the new epistemology. Epistemology does not derive from first principles, but is continuous with natural science. Second, even though naturalized epistemology does not seek to justify science, it does still examine the relation between evidence and theory. In what way? Quine notes that observation is "meager" compared to "torrential" theoretic output, that theory "transcends" available evidence: this is a restatement of his underdetermination thesis, which, like meaning holism, derives from his confirmation holism. According to that thesis, theory transcends evidence because multiple empirically equivalent theories may be equally supported by observational evidence. This relationship between evidence and theory which Quine wishes to highlight, then, is not merely the causal one. He also wishes to highlight the logical relationship—or lack thereof: there is a logical gap—between the two. Identifying causal relationships is necessary to account for how we arrive at our theory, given that we do not arrive at it from logical deduction. This identification, of course, is a project Quine had already undertaken in *Word and Object*, in which he describes the various unjustified analogical leaps by which theory is born.

Clearly, then, Quine's motivation for naturalizing epistemology is informed by the fundamental principles—particularly the indeterminacy of translation and the underdetermination of theory by evidence (both of which derive from confirmation holism)—which had long formed the core of his philosophy. It is impossible to assess his proposal without reference to this core.

Kim's non-Quinean alternatives to deductivist foundationalism

In fairness to Kim, "What is 'Naturalized Epistemology'?" does not entirely ignore the core of Quine's philosophy. But, I will argue, he does not take into explicit account its total significance.

Kim's article begins with a reminder of what he takes to be the central purpose of traditional epistemology: to outline criteria for the *justification* of our beliefs, and enumerate the extent to which our beliefs fit these criteria. These criteria must themselves be stated in "descriptive or naturalistic" terms, without recourse to further epistemic, i.e. normative or evaluative terms, lest the criteria be circular (Kim 1988, 382). Further, Kim emphasizes that the concept of "justification" is the central epistemic concept in any definition of "knowledge," and that it is an essentially normative concept which, more than any other concept, makes "knowledge" itself a normative concept. To say that a belief is justified, he thinks, is to say that it is "permissible or reasonable" to hold, and "epistemically irresponsible" to contradict it (383). To this extent Kim and Quine are in agreement: this was indeed the goal of traditional epistemology, and historically epistemologists sought to show why belief was justified in virtue of its being deducible from certain foundations.

Kim proceeds to present Quine's position. He accurately presents Quine's distinction between the conceptual and doctrinal projects. He notes in particular Quine's view that the empiricist conceptual project has failed because of meaning holism, and that the doctrinal project of deductivist foundationalism has failed because of the Humean predicament. In what follows, however, Kim does not appear to appreciate the ways in which Quine takes the conceptual and doctrinal projects to be linked, or the full significance of that linkage. This becomes apparent when Kim fails to appreciate that Quine's argument from holism against the traditional conceptual project has no serious significance for the doctrinal project:

> To be sure, Quine's argument against the possibility of conceptual reduction has a new twist: the application of his 'holism.' But his conclusion is no surprise; 'translational phenomenalism' has been moribund for many years. And, as Quine himself notes, his argument against the doctrinal reduction, the 'quest for certainty,' is only a restatement of Hume's 'skeptical' conclusions concerning induction: induction after all is not deduction….We are tempted to respond: of course we can't define physical concepts in terms of sense-data; of course observation 'underdetermines' theory. That is why observation is observation and not theory. (386)

Here Kim is clearly wrong that Quine's argument against the doctrinal project is *only* a restatement of Hume's inductive skepticism. Quine's confirmation holism not only helps to establish indeterminacy of translation (via meaning holism), but *also* helps to establish the underdetermination thesis. As I will examine in greater detail in chapter 6, confirmation holism says that hypotheses are only confirmed in conjunction with their auxiliary hypotheses, and because we are free to adjust auxiliary hypotheses at will, we may revise or hold true our hypotheses in the face of observations. This implies that our theory is underdetermined by observational evidence. Thus underdetermination is a more radical form of skepticism than Hume's: it says not only that scientific hypotheses are underdetermined by our past observations, but that they are underdetermined by *all possible* observations (Quine 1975a, 313). The problem of induction follows from the fact that there is a logical gap between statements about the past and statements about the future, or between the finite number of our observations, and the infinite scope of inductive generalizations. Underdetermination follows from the fact that there is a further logical gap between even an infinite number of observations, and the pragmatic choices that determine which auxiliary hypotheses to retain or discard, and hence the resulting theory. As we shall see quite soon, this has the effect of radically delimiting the scope of viable doctrinal projects.

In addition, the indeterminacy of translation thesis will itself have implications which further delimit the availability of alternative doctrinal projects, and this issue is independent of the success or failure of the exculpatory projects of translation Kim has in mind in the passage quoted above. As we shall see, along with indeterminacy of translation comes not only skepticism about the ability to analyze the meaning of scientific concepts, but also skepticism about the ability to analyze the meaning of philosophical concepts—including those involved in the proposal of new criteria of justification.

Because Kim isolates the conceptual project from these doctrinal implications, he believes that Quine is forcing us to make a false choice between deductivist foundationalism and naturalized epistemology. "The Cartesian project of validating science starting from the indubitable foundation of first-person psychological reports," writes Kim, "is not the whole of classical epistemology" (388). He then offers examples of the kinds of doctrinal projects that allegedly lie in between Descartes and Quine, which Quine fails to consider:

> Quine is not suggesting that we…explore others within the same framework—perhaps, to adopt some sort of "coherentist" strategy, or to require of our basic beliefs only some degree of "initial credibility" rather than Cartesian certainty, or to permit some sort of probabilistic derivation in addition to deductive derivation of non-basic knowledge, or to consider the use of special rules of evidence, like Chisholm's "principles of evidence", or to give up the search for a derivational process that transmits undiminished certainty in favor of one that can transmit diminished but still useful degrees of justification. (388)

These are, indeed, possibilities worth considering. But if Quine's principles already have some effect on doctrinal possibilities beyond deductivist foundationalism, it is worth considering whether even Kim's alternatives might also fall victim to Quinean objections. Indeed the example of Quine's rejection of Carnap's doctrinal project should already suggest that the implications of Quine's principles do extend further than the rejection of Cartesian deductivist foundationalism. But let us consider in turn each of these possibilities Kim mentions above.

Adopting some kind of probabilist epistemology does indeed seem like a natural response to the failure of the Cartesian "quest for certainty." And we can safely group together under the heading of "probabilism" the "initial credibility," "probabilistic derivation," and "diminished transmission" possibilities Kim mentions. But if there is something wrong with the first of these—if there is no way to assign initial credibility to any beliefs—there will certainly be no way for derivative beliefs to inherit any degree (diminished or otherwise) of initial credibility. It does indeed appear that because of his underdetermination thesis, Quine would say there is no objective way to assign differential degrees of initial credibility. In a recent paper, Bredo Johnsen (2005, 82–3) argues convincingly that Quine had dismissed this possibility well in advance of "Epistemology Naturalized." As early as "Two Dogmas," Quine (1953b, 45) had argued that neither physical objects nor the gods of Homer are *on any better epistemological footing* in relation to observational evidence; they differ only in their respective pragmatic values. Even though the idea that there are physical objects is a paradigm example of the kind of belief that would otherwise be taken to have "initial credibility," the underdetermination thesis does imply that one cannot assign a hypothesis greater probability than an empirically equivalent rival hypothesis. If basic beliefs cannot be invested with any objective initial credibility, non-basic beliefs cannot inherit it, or any degrees of it.

There are, of course, Bayesian probabilist critiques (among others) of the underdetermination thesis itself. The point here is not that the thesis cannot be answered, but that it must be if Quine's proposal for naturalizing epistemology is to be shown to be unmotivated. What's more, it must be answered without begging any methodological questions against the naturalist. Indeed the very language of probabilism is shot through with notions that would raise Quinean eyebrows: the notion of a *proposition* whose probability is to be measured, the degrees of *belief* in terms of which probability is formulated by many, and even the modal notion of *probability* itself, the value of which is difficult to determine extensionally.[8] Probably Quine would allow for some conversational notion of probability that expresses unwillingness to assert a sentence, but this is of little use to the justificatory project in epistemology that Kim is discussing (see Hookway (1988, 106–7)).

[8] Leaving aside questions of *degrees* of probability, notice how affirmations of probability might fall prey to the same failure of intersubstitution *salva veritate* to which other intensional statements of modality are prone. "Probably the number of planets is greater than seven" is true, but "Probably nine is greater than seven" is not true if "probably" implies less than certainty. Of course one might object that if we can knowingly intersubstitute "nine" for "the number of planets," then we are operating on the premise of certainty, not mere probability. But to the extensionalist, this objection confuses epistemology with semantics. Semantically, *whatever* number of planets there is, mathematical statements about that number it in relation to other numbers are not statements of probability, and cannot serve as arguments for probability functions.

Both traditional deductivist and probabilist proposals considered so far might have fallen into a broadly foundationalist tradition. But if underdetermination and extensionalism rule out foundationalism, what of a coherentist doctrinal project? Indeed at first Quine's confirmation holism seems to be a kind of coherentism, particularly in light of "web of belief" metaphors. But the purpose this coherentism is supposed to serve is crucial in judging its justificatory power. Quine's coherentism merely describes the structure of our beliefs: there are no initial-credibility-granting "foundations," only observation sentences that constrain (but underdetermine) the inner weavings of the web of belief. On this view, the mere coherence of a set of beliefs does not confer any logical justification on any of them. There is no question of justification apart from pragmatic justification, which is a function of the predictive and explanatory power of a theory. Perhaps a view of justification set in terms of the predictive and explanatory power of a belief *just is* a coherence theory of justification. But then Kim is wrong that coherence theory is an alternative to Quinean naturalism. Assuming that he is speaking of non-Quinean or non-naturalistic coherence theories of justification, like Bonjour's (1985), these theories require much more of beliefs than mere predictive and explanatory power: in particular, they require a plausible account of how coherent systems of belief acquire initial credibility (so as to avoid the implication that coherent fantasies might count as justified). We have already seen how Quine would rule out the possibility that beliefs derive initial credibility from observation. The typical coherentist view is to confer *a priori* justification on certain sets of beliefs. But the *a priori* is clearly not acceptable to a naturalist empiricist like Quine.[9] Of course there have been attempts to naturalize

[9] For this reason, Quine would also rule out something else mentioned by Kim: Chisholm's principles of evidence. Quine would rule out Chisholm's principles for other reasons. Chisholm seems to have posited these to account for knowledge that is reached through ways "other than the formal principles of deductive and inductive logic" (Chisholm 1977, 67). He assumes that we know most of the things we take to know, and looks for principles that would explain this. Quine would probably object that this is an unjustifiably *a priori* approach to epistemology. We could just as easily be *a priorists* about induction and deduction as the only sources of knowledge, and decide *a priori* that these principles are inadmissible, and perhaps that skepticism is true. A prominent example of one of Chisholm's principles is "Having a perception of something bein an F tends to make acceptable the proposition that something is an F" (Chisholm 1977, 68). Quine would also skeptical of the idea that observation reports are in terms of sensory qualities, and is even more skeptical of the internalist mentalism inherent in Chisholm's account.

"initial credibility" or the *a priori* through an externalist account of justification. It is to these externalist accounts that we now turn.

Both foundationalist and coherentist views of justification could be described as broadly "internalist" theories. Internalism is the more traditional theory of justification, which holds that one must have some epistemic access to or awareness of the conditions of justification. As a naturalist (and especially as a behaviorist), Quine sees no way to make sense of justification in terms of awareness. What, then, of "externalist" theories of justification, which simply abandon the epistemic access requirement? Could these still offer norms of justification without collapsing entirely into Quinean naturalism? Kim considers a number of externalistic views, which he calls "naturalistic" (but not Quinean naturalistic), stressing in particular Goldman's reliabilism and Armstrong's nomological theory. These are said to be "naturalistic" only insofar as they offer a criterion for the normative concept of "justification" in *naturalistic terms*. So these theories characterize justification by reference to factors outside of the subject's immediate awareness, such as whether their beliefs are the result of a reliable belief-forming process, or whether they stand in law-like relations to the facts that make them true. This is as opposed to Quine, who appears to dispense with the possibility of traditional epistemic justification entirely, regardless of the terms in which it described.

These externalist theories do not at first seem to suffer from any of the problems characterized by foundationalism or coherentism. However, before any of these can pass the Quinean test, we need to know more about the nature of the *methodology* each of these theories uses in order to formulate its respective criterion of justification.

In several places, Kim is careful to point out that the search for a criterion of justification does not require anything like a *definition* or an *analysis*. Early on, for example, he states that questions about the conditions of justification need not "ask for an 'analysis' or 'meaning' of the term 'justified belief'" (Kim 1988, 382). Later, he notes that (non-Quinean) epistemological naturalism "does not include (nor does it necessarily exclude) the claim that epistemic terms are definitionally reducible to naturalistic terms (398). Clearly Kim is aware of the controversy and outright naturalistic unacceptability of the notion of analyticity, and is trying to avoid the charge of begging the question by assuming non-naturalism in order to refute naturalism. Quine, after all, is famous for having rejected the notion of analyticity, in part for the same reason that led him to adopt the underdetermination and indeterminacy theses: confirmation holism.

Whether or not Kim believes that non-Quinean naturalists could offer non-analytic criteria, it remains true that the kinds of criteria he cites *were* formulated using an analytic methodology. Judging from the origin of externalist theories of knowledge in thought experiments concerning the Gettier problem (fake barns, thermometers and the like), it is arguable that Goldman and Armstrong originally used traditional conceptual analysis to formulate their criteria of justification. Even if the results of these analyses were not meant to be taken as analytic *truths*, their status as issuing from the armchair analysis of meanings, rather than as results of scientific investigation, would still seem dubious from a Quinean naturalist perspective.[10] The *terms* of the criteria offered by these externalists may be "naturalistic" in some sense, but that does not mean that the methods used to reach the criteria are naturalistic.

Of course Quine's rejection of analyticity is not entirely uncontroversial. Recent years, in particular, have seen the resurrection of the case for analyticity from the perspective of conceptual role or two-dimensionalist semantics. (In a later book (2005), Kim joins the trend.) But the new approach is still probably more controversial than Quine's original critique: as we shall see in chapter 3, it relies crucially on semantic and modal concepts that Quine's naturalism would call into question. So it is important that Kim (at least at this stage) *is* careful not to imply that the only philosophic criteria are analytic definitions. In the final section of his paper, he does discuss a kind of philosophic criterion which appears, at first, not to have any obviously non-naturalistic commitments: epistemic

[10] In chapter 3, we will examine more of the ways in which the methodology of conceptual analysis conflicts with more basic tenets of naturalism (particularly because Kim later comes out in favor of a limited version of conceptual analysis).

supervenience (Kim 1988, 399–400). In my next section, I will show that even the supervenience relation, as ontologically minimalist as it seems, is challenged by the same Quinean problems that cast doubt on analyticity.

Kim's alternative methodology: epistemic supervenience

Few have done more than Kim to explain the idea of supervenience, particularly in philosophy of mind. I do not here wish to evaluate the overall adequacy of appeals to supervenience as a tactic in philosophy, whether in epistemology or any other discipline. I simply want to examine whether appeals to supervenience are consistent with a Quinean naturalistic outlook. Kim himself avoids appealing to conceptual analysis, presumably to avoid begging the question against Quine. If, for some reason, the appeal to supervenience also turns out to be inconsistent with the same principles of Quinean naturalism that motivate the case against analyticity, then it seems that invoking it would *also* beg the question. Supervenience would be at least as controversial as analyticity, and therefore no panacea as an alternative to Quinean naturalism. Of course, Quine's fundamental naturalistic principles may be wrong, in which case supervenience (and also analyticity) could be acceptable. But Kim does not attack Quine's fundamental principles, so it is imperative to determine whether they are in fact compatible with the notion of supervenience.

Appealing to supervenience is a widely accepted method not only in philosophy of mind but also in ethics (wherein, arguably, the concept was first developed). It helps to describe the connection between a higher level property (like a mental or ethical one) and a lower level property (like a physical one), by means of stating that there can be no difference in the higher level property without a corresponding difference in the lower. It is natural to think of epistemic properties like justification as higher level, and thus propose that they too supervene on natural properties, and that epistemological criteria identify these supervenience relations. Kim (1988, 399) states his understanding of epistemic supervenience loosely as follows:

> [I]f a belief is justified, that must be so *because* it has certain factual, nonepistemic properties, such as perhaps that it is 'indubitable,' that it is seen to be entailed by another belief that is independently justified, that it is appropriately caused by perceptual experience, or whatever. That it is a justified belief cannot be a brute fundamental fact unrelated to the kind of belief it is. There must be a *reason* for it, and this reason must be grounded in the factual descriptive properties of that particular belief.

Before getting into the details of the supervenience proposal, it is important to point out that any claim that an appeal to supervenience could be "naturalistic" should be greeted with *prima facie* skepticism. After all, the idea of supervenience was originally developed by ethical *non*-naturalists like G.E. Moore to characterize their view of the relationship between value properties and natural properties. Moore would have agreed that value properties supervene on the natural, but would have insisted that they themselves are *non*-natural. This gives us reason to think that whatever sense of "naturalism" someone might invoke to treat supervenience relationships as "naturalistic" is idiosyncratic, at best. As with our discussion of conceptual analysis, it is not enough that properties supervene on natural properties. If we must use some non-naturalistic methodology to discover or describe this supervenience, the naturalistic status of the supervenience base is irrelevant.

J.P. Moreland (1988, 35–57) has argued that supervenience is not a notion that naturalists like Quine would find acceptable. Briefly I will describe his case and then consider potential objections to it. Moreland notes two important concepts of supervenience, weak and strong, and argues that only the second is strong enough to serve the naturalist's purposes. Property A *weakly supervenes* on naturalistic property B if and only if objects whose B-properties are indiscernible are also A-indiscernible. If A is an epistemic property like justification, then this would mean that two subjects identical in subvenient naturalistic B properties will not have beliefs differing in their justification. Moreland notes that this concept of supervenience is not strong enough to give the naturalists what they want, in effect because it is consistent with accidental (but systematic) correlations between basic and supervenient properties. This version of supervenience does not require, for example, that natural properties actually *determine* epistemic ones. One advocate of a naturalistic supervenience view, Daniel Bonevac (2001, 159) agrees on this point, noting that weak supervenience allows "no cross-world inferences" of the sort needed for naturalistic determination.

Closer to what a naturalist needs to explain epistemic properties by reference to naturalistic ones is *strong supervenience*. Property A *strongly supervenes* on property B if and only if *necessarily*, objects whose B properties are indiscernible are also A-indiscernible. If A is an epistemic property like justification, this would mean that any two subjects with identical subvenient natural B properties *must* not have beliefs differing in their justification. The main question to ask is what sense a naturalist could make of the idea of "necessity." Conceptual necessity—truths necessary because they derive from the meaning of concepts—would seem to be inadmissible for the usual reasons associated with the rejection of analyticity. Quine is also generally skeptical of *any* substantive modal notions, such as necessity, because statements of necessity are referentially opaque in the same way that statements reporting beliefs can be. According to Quine, both fail the test of intersubstitutivity *salve veritate*: "Necessarily, the number of planets is greater than seven" would be false according to modal logic, but not "Necessarily, nine is greater than seven." More generally, Quine regards with suspicion any intensional notions, whether belief or necessity, on the grounds that they do not have the crisp identity conditions associated with extensional logic. Logic understood extensionally, Quine (2004a, 335) thinks, is "the grammar of strictly scientific theory." So Quine's extensionalism itself falls out of his more general naturalism—and overrules a methodology of supervenience grounded on conceptual necessity. And let us not forget that extensionalism itself plays a key role in Quine's argument against analyticity: unlike reference, which can be described in extensional terms, meaning—and truth in virtue of meaning—is putatively intensional and in need of special explanation.

Might another more acceptable notion of necessity be available? Moreland considers two remaining notions: metaphysical (Kripkean) necessity and the nomological necessity. Metaphysical necessity is itself an expression of the post-Quinean revolution in modal logic that seems, in many ways, to obviate Quine's critique of modality by providing a powerful, extensional explication of modality, owing to its model-theoretic quantification over possible worlds. A property is metaphysically necessary if it obtains in every possible world. But as Robert Brandom (2001, 598) points out, this new expressive power does not settle the question between advocates and naturalistic critics of modality:

> The Kripke semantics is not even a candidate for providing such a reduction, because it owes its extensional character to the introduction of new primitive notions, possible worlds and accessibility relations . . . that are themselves richly modal (and whose deployment turns out to require further metaphysically nontrivial commitments concerning about what is essential and what accidental). Any legitimate reasons to doubt the legitimacy of talk of necessity and possibility are just going to be passed off and transformed into corresponding reasons to doubt the legitimacy of appeal to such primitives.

Of course it is not that Quinean naturalists would reject possible worlds just because of their status as abstract objects, accessible by *a priori* intuition. Quine accepts the existence of mathematical objects as an ontological commitment generated by science's reliance on mathematics. One question that Moreland does not consider is that Quine might take the acceptance of modal objects (i.e., possible worlds) as practically indispensable for science as well, and on this basis permit them acceptable for use in describing supervenience relations. This seems to be the approach of Chalmers (1996, 66), but Chalmers does not explain why possible worlds should be treated as indispensable primaries while other putatively intensional concepts (e.g. of the mental) need to be naturalized. There are more options available to those who would treat modal notions as pragmatically indispensable, but we will wait until chapter 3 to discuss them in more detail.

Perhaps it is best to turn away from metaphysical necessity—which was never likely to satisfy the naturalist in the first place—and turn to nomological necessity. To say that some connection between properties is nomologically necessary is just to say that the connection is covered by some law of nature. Of course to exhibit nomological necessity, it is not enough that the supervenient and subvenient properties regularly correlate. As Moreland points out, some further *explanation* must be offered showing that the laws of physics do actually guarantee that the subvenient determines the supervenient. It is the naturalist's burden of proof to show that this explanation, together with the relevant physical laws, exists.[11]

Of course this criticism leaves open the possibility that naturalists might someday shoulder that burden, and supply the relevant laws, or at least learn a way of outlining them without stating them in detail. To see what this would involve, let's consider the parallel question of supervenience of the mental. Kim argues in an earlier essay that for mental properties to supervene requires that there be certain psychophysical laws (Kim 1985). This, interestingly, is a view that he developed in opposition to Davidson's (1984) "anomalous monism," a view which argues *in principle* against the possibility of such nomological necessity of the mental, and which is motivated by some of naturalism's favorite principles, such as the indeterminacy of meaning (62, 313).[12] Kim recognizes that Quine would take the indeterminacy thesis as refuting supervenience claims, and contends that by defending the existence of psychophysical laws in virtue of which supervenience relationships would acquire their necessity, he could, in effect, refute the indeterminacy thesis. This is the closest Kim comes to challenging one of Quine's fundamental principles—indeterminacy—though not by examining its roots in underdetermination and confirmation holism (which we shall do in chapter 6). Nevertheless, if Kim could supply and defend the relevant psychophysical laws, he would confront indeterminacy directly and reject at least one of naturalism's basic tenets in a non-question-begging manner.

Unfortunately for Kim, the sorts of psychophysical laws he suggests are of a controversially explanatory character to begin with. Traditionally, the possibility of psychophysical laws has been considered challenging because of the multiple realizability of the mental: it is hard to see how a given mental type is reducible to or identifiable with only one physical type or natural kind. Kim's 1985 solution to this problem is to argue that psychophysical laws might be formulated as long disjunctions specifying the many different physical types that could realize mental properties. Of course Kim admits we may never be able to articulate all of the disjuncts, but he is satisfied to require that they

---

[11] See, for example, Vahid (2004, 15).
[12] It is telling that the later Quine himself endorsed anomalous monism.

merely be formally representable in theory. But I think this conception of law already poses preliminary problems for the possibility of the nomological necessity of the mental, since normal science does not seem to use such unspecified disjunctions as laws. These disjunctive "laws" lack law-like character. As Stephen Daniel notes, they don't explain psychophysical covariation so much as "catalogue" it; they state merely sufficient conditions for the mental (Daniel 1999, 231–2).[13]

It is, of course, impossible in this chapter to wade into the debate about what makes for genuine lawlikeness in scientific explanation. But it should suffice to say that Kim faced an uphill battle in defending the idea that his proposed psychophysical laws are even laws, to say nothing of whether or not they really exist. This makes it implausible for him to assert that one can simply choose between the indeterminacy thesis and his view of psychophysical laws: his view would need to be more obviously true than indeterminacy for that to be the case, and it is not. And since Kim has not addressed so many of the other positions that lay a foundation for the indeterminacy thesis (such as confirmation holism), his argument appears to amount to a kind of special pleading. Kim himself (2005) later appears to abandon the nomological account of necessity, in favor of a conceptual account which he thinks offers more explanatory power. But while we are here looking for an account of necessity to underpin descriptions of supervenience relations, the later Kim appeals to supervenience relations to underpin conceptual necessity. So this doesn't accomplish what we are looking for.

Assuming that both metaphysical and nomological conceptions of necessity are not adequate to provide a naturalistic account of strong supervenience, we should consider whether there is any conception apart from necessity that could account for the element of dependence in the strong supervenience relation. One such conception is discussed by Daniel Bonevac (2001). Bonevac proposes, in essence, an epistemic account of supervenience: to say that one property supervenes on

[13] Daniel even argues that even without presupposing the radical indeterminacy of the mental, Kim's laws run into trouble under the assumption of *limited* mental indeterminacy. If there are just a few mental states that are indeterminate, it seems a genuine psychophysical law would need to identify the supervenience base that explains why certain mental states are determinate and others are not. And, Daniel thinks, this is in principle impossible, because it is in the nature of the indeterminate that there is never a sharp line between it and the determinate. See Daniel (1999, 233-4).

another is just to say that the existence of the supervenient property can be justified by appeal to a "defeasibly sufficient reason," a premise concerning a naturalistic fact made relevant by a "fainthearted conditional" (2001, 153). A fainthearted conditional is one stating a relationship between antecedent and consequent "provided that conditions are suitable," in other words a conditional with a *ceteris paribus* clause. This account then turns on recognizing the truth conditions of the fainthearted conditional.

In response to Bonevac's proposal, I will not even comment on the difficulty of using an epistemic concept of supervience in order to support a concept of epistemic supervenience: if we already know what defeasibly sufficient justification is, we have already settled a great deal of traditional epistemological inquiry. This aside, it is a more immediate concern that the truth conditions Bonevac invokes for fainthearted conditions are to be spelled out in modal terms: "'If *p*, then (provided that conditions are suitable) *q*' is true at world *w* iff *q* is true in all *p*-worlds in which conditions are suitable for assessing (relative to *w*) what happens when p" (Bonevac 2001, 149). We are then faced with the same problem identified by Brandom above in relation to metaphysical necessity: reducing it requires further, naturalistically controversial modal notions. Now Bonevac also invokes David Lewis' conception of "Humean supervenience," the proposal that the modal itself might be made to supervene on the non-modal. But if, as I have urged, supervenience itself requires modal notions, then it is almost incoherent to think about how the modal itself could be explained modally in terms of the non-modal.

Last of all, let us consider the question of the naturalistic acceptability of the proposed supervenience *base* for epistemic properties. Vahid 2004, 168) has argued that epistemic supervenience may be easier to establish than mental supervenience, given that epistemic properties may not be subject to multiple realizability in the way that mental ones are said to be. This, however, presupposes that being justified is not itself some kind of mental state, or somehow dependent on mental states. It is not entirely clear that we should accept this presupposition. Consider, for example, that a belief is very often justified by reference to other beliefs (some maintain that it is *always* by reference to other beliefs). If these other beliefs, qua mental states, could themselves be multiply realized, it would seem to follow that the property of being justified in terms of these beliefs would also be multiply realized. Of course justification is a relation among beliefs, not necessarily a belief itself. And relations are not reducible to their relata. But surely relations would at least *supervene* on relata if anything. So at best it is an open question as to whether justification is multiply realizable. Even if it is not, and epistemic supervenience does not face the same barrier as mental supervenience, there is still a heavy burden of proof on the advocate of epistemic supervenience to supply some law, disjunctive or otherwise, that goes beyond the mere identification of a mere property covariance.

Apart from the fact that justification may be multiply realizable in a way that makes it difficult to supervene nomologically on specific properties, there is the fact that those properties—in particular, beliefs—may face naturalistic problems of their own. Kim accepts fairly uncritically that the naturalized epistemologist wants to study the relationship between perception and belief. In the next section, however, I will show that Quine was not this kind of naturalized epistemologist—and for reasons that derive from the usual principles of his system.

None of the above section is meant, of course, to challenge the use of supervenience in philosophic criteria *per se*. Rather, it is to show how deeply divided naturalist and non-naturalists are from each other, down to the relevance of basic methodological concepts like supervenience.

Quinean doubts about supervenience on beliefs

To embelish his point that any naturalized epistemology dispensing with normativity must abandon genuine epistemology, Kim argues that even the concept "belief" is normative. Invoking Davidson (1984), he argues that belief attribution involves "radical interpretation," which requires an interpretive theory viewing the total system of the subject's beliefs as "largely and essentially rational and coherent" (Kim 1988, 393). If we could not view other beings as largely rational, we could not even identify them as "cognizers" or as having beliefs in the first place. Therefore, abandoning normativity also means abandoning the concept "belief." This, Kim says, is problematic for naturalism, which seeks to "identify, and individuate the input and output of cognizers" (392).

The trouble is with Kim's assumption that Quine's naturalized epistemology involves the assignment of *individually meaningful beliefs* to subjects in order to examine the cognitive input/output relationship. But Quine is famous for arguing that interpretation is really a domestic version of "radical translation," and that translation is indeterminate. Translation is indeterminate, for Quine, for the reasons we have already considered. It is only the use of pragmatically-generated *normativity*-imbued "analytical hypotheses" that permit translation to move forward, but not in a determinate way. So, on Quine's views, "belief" attribution does indeed require a certain kind of normativity, and it is for this very reason that it is indeterminate, which is to say that there are no objective facts of the matter—no naturalistically respectable facts—determining our assignment of beliefs. This would seem to cast doubt on the possibility that beliefs might form the naturalistic supervenience base for epistemic properties.

But if Quine is not interested in objective beliefs, what kind of "output" does his naturalized epistemology deal with? At one point, Kim actually considers that Quine might simply refuse to treat the cognitive outputs of interest as "beliefs" at all (Kim 1988, 394). Kim considers that Quine may consider various neural states, instead of beliefs, to be the appropriate sorts of output, but responds that to identify the appropriate neural states would still require pairing them with interpreted beliefs. Quine's only recourse would be to bypass anything even remotely related to normatively interpreted beliefs, a consequence Kim takes to be unacceptable for anyone doing anything like epistemology.

Quine may not find Kim's attempted *reductio* so absurd. Kim misses another kind of "output" that is near and dear to Quine's heart: language. That this is the intended subject of study in his version of naturalistic epistemology is made quite clear in *The Roots of Reference*(1974, 35):

> We want to know how men can have achieved the conjectures and abstractions that go into scientific theory. How can we pursue such an inquiry while talking of external things to the exclusion of ideas and concepts? There is a way: we can talk of language. We can talk of concrete men and their concrete noises. Ideas are as may be, but the words are out where we can see and hear them. And scientific theories, however speculative and however abstract, are in words.

As I have already suggested, instead of evaluating our "outputs" in the traditional way. Quine's proposal is to examine how they were generated and came to serve our pragmatic purpose, *given* that they were not generated by being logically justified.

Avoiding reference to beliefs is, of course, the essence of Quine's methodological behaviorism. Quine's behaviorism is itself a consequence of his more fundamental principles, his extensionalism and his indeterminacy thesis, and more generally, his naturalism. Understanding this, we are also in a position to see that the range of possible naturalistically permissible supervenience bases has been delimited quite narrowly. Even if justification did not face problems of multiple realizability, it now seems that any traditional theory of epistemology making reference to beliefs (whether via their coherence or the reliability of their formation) is a non-starter for Quine.

Conclusion

Although I have not focused much on Kim's normativity argument in the above, it is worth observing how eliminating Kim's alternatives reflects upon that argument. Kim has argued that if one dispenses with logical normativity in the traditional sense, one also dispenses with the normativity of the concept "knowledge"—and therewith, the normativity of the discipline of epistemology itself. But with the core of Quine's philosophy in mind, it should come as no surprise that Quine talks as little about "knowledge" as he does about "belief" in "Epistemology Naturalized." As long as his epistemology can concern itself with examining the relationship between evidence and language, Quine thinks it still has an important task to accomplish, even if this involves no reference to "knowledge" or "justification." Quine's naturalism, after all, is pragmatist. He is not concerned with preserving the *traditional* concepts of "justification," "knowledge," or even "epistemology." Here it is useful to note Quine's (1981a, 474) reply to Barry Stroud, who also insists that naturalized epistemology does not seem to achieve the goals of traditional epistemology:

> Stroud finds difficulty in reconciling my naturalistic stance with my concern with how we gain our knowledge of the world. We may stimulate a psychological subject and compare his resulting beliefs with the facts as we know them; this much Stroud grants, but he demurs at our projecting ourselves into the subject's place, since we no longer have the independent facts to compare with. My answer is that this projection must be seen not transcendentally but as a routine matter of analogies and causal hypotheses. True, we must hedge the perhaps too stringent connotations of the verb 'know'; but such is fallibilism.

Having repudiated conceptual analysis, Quine is not, in general, interested in preserving any traditional concepts. What he hopes for, at most, is to *explicate* our concepts: to take those which already serve some clear, practical purpose, and clarify and/or modify them as needed. In chapter 5, we will see how Quine proposes to explicate the concept of "justification" in pragmatic terms. In this case, he finds little about the traditional concepts of epistemology to be usable.

We may disagree with Quine that there is so little to salvage from traditional epistemology. But if we are right, it is our responsibility to show where Quine has gone wrong. I have argued that Quine's objections to traditional epistemology run deep: they grow naturally from the fundamental principles of his naturalism, i.e., from his underdetermination thesis, his indeterminacy thesis, and his extensionalism. Kim's critique of Quine is severely limited because he fails to consider the numerous ways in these principles delimit the scope of viable epistemological alternatives, rendering naturalized epistemology the only going option. Even his apparently independent objection from the normativity of the concept "belief" fails for the same reason. In short, Kim has given us no reason to think that Quine's epistemology is unmotivated, or that it is somehow internally inconsistent.

How then should we critique Quine? I recommend challenging him at the root. Challenge the underdetermination thesis which Kim concedes, but which also underlies the powerful indeterminacy thesis. Propose alternatives to confirmation holism, which underlies both indeterminacy and underdetermination. Challenge his extensionalism with examples of intensional concepts that appear to be indispensable to scientific discourse. Even reconsider the basic thesis of naturalism, the idea that science—rather than first-handed commonsense observation—is our only source of knowledge. Most of these challenges will be made in chapter 6. But before that, we need to examine in more detail whether Quinean behaviorism is completely justified in rejecting the concept "belief."

CHAPTER 3

NATURALIZING BELIEF FOR NATURALIZED EPISTEMOLOGY

In my first chapter, I outlined a new taxonomy of naturalized epistemologies, divided according to the methodological concepts exploited by varieties of “conceptual projects” used by various naturalisms. I argued that “optimistic” naturalized epistemologists might naturalize knowledge by analyzing the *concept* “knowledge” in accordance with different semantic theories (e.g., analytic or two-factor), or by exploring the metaphysics of the *fact* of knowledge by way of metaphysical identity criteria (e.g., supervenience). I noted in passing, however, that the naturalistic acceptability of many of these methodological concepts could easily be called into question. In this chapter, I want to explore and defend that claim in greater detail. But instead of showing how these methodological concepts would be applied to the concept “knowledge,” I want instead to survey their application to one of the concepts in terms of which “knowledge” is usually analyzed: the concept of “belief.”

There are several reasons for using “belief” as the case study for the naturalistic acceptability of these methodological concepts. First of all, it is not at all uncommon for one part of the *analysans* of “knowledge” to play proxy in debates over the naturalizability of knowledge. More typically, epistemologists will focus on the naturalizabiltiy of “justification,” particularly its normative element. In my first chapter, I surveyed some of the debate concerning the naturalizability of the normative, and concluded that naturalists had a stronger case for naturalizing the normative than is usually conceded—provided that they are permitted the usual range of naturalization methodologies we are currently calling into question. So one reason to explore other elements of the “justified true belief” complex is that “justification” has already been examined in some detail. It had been examined because normativity was thought to be a most distinctively non-naturalistic property, and if naturalism could not countenance it, surely epistemology could not countenance naturalism. Another reason to examine “belief” is that it is directly connected to an even greater collection of properties thought to be distinctively non-naturalistic: both inten*t*ionality and inten*s*ionality (as well as a sizeable normative element of its own). So “belief” is as much of a challenge to naturalizing knowledge as “justification,” if not more so.

There are, of course, probably even more proposals for “naturalizing” belief and intentionality than there are for naturalizing normativity. It is, after all, the subject matter of great swaths of the philosophy of mind. The varieties of naturalization proposals for belief basically parallel those we examined for naturalizing knowledge, but in this case questions about the subject matter to be naturalized are in many cases *the same* as the questions about the methodological concepts that will do the naturalizing. Notice, for example, how many of the naturalization proposals we have examined revolve around different theories of *reference* (because we are concerned with the reference of the concept “knowledge”). But theories of reference are of interest to naturalizing belief not only because we want to know more about the reference of the *concept* “belief,” but because the *fact* of belief itself seems to involve an intimate connection to the fact of reference: beliefs are thought to be individuated by their content, and *content* is often thought to be a function of reference (among other things). As we shall see at the end of the chapter, even if the more consistent naturalization proposals for “belief” can dispense with objections concerning putatively non-naturalistic methodological concepts, questions will linger concerning their treatment of their subject matter, particularly insofar as theory of reference is needed for understanding the content of scientific beliefs, which are of special concern to naturalized *epistemology*. This is the second reason it is useful to examine the “belief” component of the traditional analysis of “knowledge.”

So if it should turn out that we cannot naturalize a concept of “belief” usable by the naturalized epistemologist, this inability may count as an objection to naturalized epistemology itself. Insofar as the naturalizability of belief is tied up with naturalizing *any* phenomenon by way of a “conceptual project,” understanding why we cannot naturalize belief in this manner may also end up reflecting back on questions about the naturalization of normativity and of justification. Understanding this may, in other words, feature a more fundamental objection to naturalized epistemology than some of the traditional critiques.

In this chapter, therefore, I will examine naturalization proposals using a taxonomy similar to the one developed in my first chapter. This time, however, I will develop in greater detail the naturalistic objections to various methodological concepts. I will first examine "analytic naturalism" about belief, and develop the objection that naturalization of this variety is unacceptable to the naturalist because of its reliance on the method of *a priori* conceptual analysis (in spite of some recent defenses of this method against traditional Quinean objections). I will then consider "conceptually regulated scientific naturalism" which relies in part on conceptual analysis, and in part on other factors to determine reference. I will argue that this second proposal is non-naturalistic, not only because of its appeal to conceptual analysis, but also because of its reliance on intensional concepts at odds with basic naturalistic precepts. Finally, I will examine "conceptually indifferent scientific naturalism", a naturalization proposal that does away with conceptual analysis entirely. I take it that the final version is the most naturalistically palatable, and for this reason my polemics against *it* will be the most important in the chapter: I will argue that this final version, while generally free of non-naturalistic methodological concepts, fails to furnish us with a concept of "belief" *usable for epistemological purposes*. Having surveyed these alternatives, I will conclude that no obvious candidate for naturalizing belief is available for what I have labeled "optimistic" naturalized epistemology.[14] As a result, the only consistent form of naturalism possible seems to be of the pessimistic variety.

Before I come to any conclusions about the viability of *any* belief-naturalization proposals, I need to explain why it is that any naturalists feel it is important for their epistemological purposes to naturalize belief in the first place. This is, of course, a point that the pessimists will disavow, but it is

[14] In my first chapter, I counted "supervenience" naturalism in a separate "metaphysical" category apart from proposals based on theories of reference. Between that discussion and the more detailed discussion of Kim's views on supervenience in the second chapter, it should be clear that supervenience is of little use for naturalization purposes on its own, that it usually must be conjoined with reference-theoretic concerns. For this reason we will not treat it separately this time, but in conjunction with conceptually-regulated scientific naturalism, a spot where Kim himself has most recently placed it (Kim 2005).

important to be precise about *what* they are disavowing. In the first section of this chapter, therefore, I will present the claims by "optimistic" naturalists such as Kim, Goldman, Kitcher and Kornblith regarding the need to naturalize belief

Why naturalized epistemology needs naturalized beliefs

Do naturalized epistemologists recognize that their project depends on a naturalistic account of belief? Consulting the relevant literature reveals that they do.

Writing in the critique of Quine that we have now examined in some detail, Jaegwon Kim argues that even if Quine's doctrine fails to support a concept of epistemic justification, it may share enough in common with traditional epistemology to warrant the title of "epistemology" if it shares a concern for the subject matter of the formation of beliefs. This is because, as Kim suggests (1988, 392), even a naturalized epistemology would need to:

> . . .identify, and individuate, the input and output of cognizers. The input, for Quine, consists of physical events . . . and the output is said to be a "theory" or "picture of the world"—that is, a set of "representations" of the cognizers environment. . . . In order to study the sensory input-cognitive output relations for the given cognizer, therefore, we must find out what "representations" he has formed as a result of the particular stimulations that have been applied to his sensory transducers. Setting aside the jargon, what we need to be able to do is to attribute *beliefs*, and other contentful intentional states, to the cognizer.

Kim raises the issue in order to argue that "belief" itself is an inherently normative concept, and that any difficulty naturalism has with normativity will translate into a difficulty with "belief." In my first chapter, I argued that naturalism's difficulty with normativity is not as obvious as might seem to some. Even if that is true, however, Kim's point is still significant: it suggests that if there is *some* difficulty naturalists have with "belief" (because of its normativity or some other concern), naturalists may have difficulty naturalizing epistemology. He does not, however, say much more here about why he thinks contentful intentional states are so important to epistemology.

Kim does have much to say, of course, about how beliefs might be naturalized. He has spent much time developing the notion of "supervenience," and argues in numerous places (especially Kim

2005) that intentional content must supervene on the physical, by way of a reductive explanation drawing on functional conceptual analysis. We will discuss his views of naturalization in detail under the second naturalization proposal below (conceptually-regulated scientific naturalism).

Other naturalized epistemologists say more than Kim about why naturalized epistemology needs to mention beliefs. One example is Alvin Goldman, whom we have already described as an "optimistic" naturalist of the analytic variety. Goldman (1986, 162) explains why his epistemology assumes "the existence of beliefs and other propositional attitudes":

> To say that something has content is to say that it has semantic properties: meaning, reference or truth-conditions, for example. Given my epistemological perspective, truth-conditions are especially important. Unless mental states have truth-conditions, there can be no true or false beliefs; *a fortiori* there can be no mental processes with epistemological properties that interest us, such as, power and reliability. My investigation of such properties of mental processes could not get started: it would be devoid of relevant subject matter.

Goldman goes on to say that he thinks any epistemology—naturalistic or otherwise—must find a place for mental content. If an epistemology wishes to take knowledge seriously, it must take belief seriously. Even a form of pragmatist epistemology unconcerned with knowledge would need to talk about agreement and disagreement, which would still presuppose semantic content (162–3). Goldman recognizes, of course, that Quine's naturalized epistemology or other "speech-act" epistemologies focus on explaining what people say, not on "internal ideas and beliefs" 163).[15] But since Goldman is an "optimistic" naturalist, he sets these latter naturalisms aside. Instead of considering epistemologies that do away with mental content, Goldman examines some of the popular naturalistic criticisms of mental content. I will examine a portion of his response here, as it serves as useful set-up for the naturalization proposals we are about to consider.

Goldman first considers the eliminativist arguments of Paul Churchland (1981; 1996), which stem from treating content attributions as expressed by a kind of theoretical folk psychology. If folk psychology has limited explanatory or predictive power, it may prove to be false, and the entities to

[15] We could easily think of Williams's deflationism as an example of a "speech-act" epistemology.

which it refers (presumably, defined by their theoretical role) would not exist. This seems particularly likely to Churchland, since he thinks the entities of folk psychology are irreducible to neuroscience—particularly entities constituted by some kind of "language of thought."

Goldman (1986, 164–7) responds to Churchland with three rejoinders. First, he wonders if folk psychology should be expected to be able to explain everything Churchland says it fails to explain (mental illness, creative imagination, intelligence differences, ball-catching abilities, or the functions of sleep). He also challenges Churchland's presupposition that contentful states would need to be composed of an internal mental language. I agree with both of Goldman's criticisms here, and will not pursue the issue further. Second, Goldman wonders whether folk psychology need be understood as a kind of theory, and mentions problems philosophers have had defining the functionalist program in philosophy of mind. In his later work, Goldman (1995) will of course articulate a positive alternative to the "theory-theory" approach to folk psychology: the so-called "simulation" approach. I will examine this proposal in chapter 4.[16] Finally, Goldman argues that even if folk psychology is theoretical, its ontology need not be rejected just because its predictive/explanatory power is not perfect. He concedes that we will probably never produce a "strong" reduction of folk psychological types to neuroscientific types, but suggests that other forms of reduction may be possible. In particular, he thinks a corrected version of folk psychology might reduce somehow to properties at higher levels of analysis, as other concepts in the special sciences might.

There is a tension between this response and his second point, of course, because usually philosophers who propose reductions to higher-level properties have *functional* properties in mind. Given his discontent with existing functional analyses of "belief," and also given the need to preserve folk psychological ontology in the face of its possible theoretical limitations, Goldman might appreciate our second naturalization proposal to be discussed below (conceptually regulated scientific

[16] It turns out that Goldman's version of simulation theory doesn't work to naturalize belief, even if it accomplishes other purposes (a point Goldman himself concedes). Other versions of simulation theory better suited for naturalizing belief end up to be incompatible with important scientific evidence, but we will settle this later.

naturalism), which distinguishes between reduction and "reductive explanation." It derives from a theory of reference designed to accommodate the fact that we often refer to more or different properties of a phenomenon than we can predict in advance.

Goldman then considers a second set of objections to mental content raised by Stephen Stich (1983). Rather than arguing that folk psychological properties are irreducible to neuroscientific properties, Stich focuses instead on irreducibility to cognitive science. He alleges that folk psychological attribution depends too much on the attributor's theory of the world, and risks illegitimately disenfranchising too many other subjects (who happen to have different theories) as genuine cognizers. (For example, "Mrs. T" who suffers from memory loss can tell us "McKinley was assassinated" but cannot tell us anything else about McKinley or assassination. We feel we cannot assign genuine content here if there is nothing else she can say.)

Goldman (1986, 167–9) responds to Stich by questioning whether we should expect sentences of English to capture the entire content of a subject's belief. He suggests that we might understand content as a function of ranges of possible worlds envisioned by the content-holder, and further mentions externalist theories of content (such as Burge (1979)) that would permit assigning content based on subjects' background beliefs, rather than simply their utterances. Now as it turns out, both the possible worlds and externalist views of content are given ample consideration in the second belief naturalization proposal below ("conceptually regulated scientific naturalism"). So the viability of Goldman's proposals will in effect be examined there.

Whether we consider his responses to Churchland or Stich, Goldman is clearly only leaving us promissory notes, entitling us to belief-naturalization once someone else's theory has been vindicated. Elsewhere (1986, 16), he writes that cognitive science may be able to do without a positive account of mental content. But his insistence on showing how such an account is at least *possible* (with the arguments above), shows that naturalized epistemology, unlike cognitive science, may not be able to sustain itself without this account. His reliabilist approach is clearly dependent on making sense of the reliability of *belief*-formation processes. This suggests that his "optimistic" naturalized epistemology will default if his promissory note cannot be redeemed.

Another naturalized epistemologist who explicitly considers the question of the naturalizability of belief is Philip Kitcher (1992). Kitcher also considers objections to mental content from both Churchland and Stich, but focuses specifically on their significance for "the *current* practice of naturalistic epistemology and philosophy of science." He acknowledges that folk psychology may someday be called into question by advances in neuroscientific or cognitive psychological research, but that until that day comes, there seems to be little for the naturalized epistemologist to do except exploit the language of mental content rather than a previously unknown advanced scientific language:

> [I]n advance of developing this [scientific] language in sufficient detail to account for the sophisticated reasoning that appears to occur in human inquiry, there is no way of formulating naturalistic claims about cognitively optimal strategies. The very advantage on which eliminativists sometimes insist—to wit, the display of kinship between human beings and other cognizers—is also a bar to the adumbration of naturalistic epistemology along eliminativist lines. For the goal of naturalistic epistemology and philosophy of science is to understand and improve our most sophisticated performances, and about these eliminativists have presently very little to say.
>
> If we accepted the eliminativist indictment of traditional propositional approaches to cognition, then prospects for naturalism would be discouraging. In effect, we would be confronted with the choice between an inadequate idiom and one not yet developed (1992, 109).

Kitcher goes on to suggest that we may not need to make this choice, because the "eliminativist indictment" is often overstated. Clearly this is true, if Goldman's preceding objections are any indication. Kitcher mentions that while some cognitive research clearly shows that much scientific cognition is not propositional, this research still retains many of the "categories of traditional epistemology." Interestingly, one of the sources he cites is none other than Goldman's *Epistemology and Cognition,* which we have found not to contain any new insights on the subject. Kitcher (1992, 110) closes by suggesting that for the time being, we should use an approach that combines two options:

> (i) aim to develop the preferred rival idiom and defer projects of epistemic appraisal until they can be reformulated in these terms, and (ii) continue to use whatever

> resources from empirical studies of cognition can be used to formulate and address normative epistemological enterprises.

I take it that what Kitcher means by combining these approaches is to work on developing the "preferred idiom" *without* deferring "projects of epistemic appraisal," for otherwise one could not also continue to use existing resources to address the normative questions he mentions.

Like Goldman, then, Kitcher seems to be betting on the outcome of someone else's naturalization projects. Since we know that he owes no heavy debts to folk intuitions, it seems likely that rather than favoring the second approach below ("conceptually regulated scientific naturalism"), he may instead favor the third ("conceptually indifferent scientific naturalism"). This would make sense, particularly since Kitcher's own two-factor theory of reference is the general type of theory of reference that undergirds naturalization proposals of the third type. Not only would such an approach be useful for determining the reference of "belief," but insofar as theories of content themselves are intimately linked to theories of reference, Kitcher's theory may itself play a role in explaining the nature of content and thus, of belief. (At the end of our discussion of the third approach, we will, in fact, briefly mention how Kitcher's theory might be used to articulate a notion of content that is compatible with understanding the content of our "most sophisticated performances," i.e. our advanced scientific beliefs.)

A final naturalized epistemologist who helps to indicate the centrality of a naturalistic account of belief to his epistemology is Kornblith (2002). In Kornblith's view, knowledge is a natural kind, to be understood according to the causal homeostatis theory of reference of Boyd (1991), a two-factor theory similar in many regards to Kitcher's. As discussed in our introductory chapter, Kornblith argues that the category of *knowledge* has "theoretical unity," because of the way in which this particular capacity of organisms reliably produces true beliefs and permits successful action. One might think that Kornblith would, therefore, seek to show *belief* to be a natural kind, but this is not entirely clear. What is clear is that he does seek to offer some manner of naturalistic account of belief, drawing on investigations in cognitive ethology (the study of animal cognition). He argues that we can understand beliefs as a particular sort of information-bearing state, where information-bearing states are understood as internal "representations" that enable everything from thermostats to ants to process information from the environment and respond to it in a particular fashion (2002, 35–7). What distinguishes belief from mere informational state is that belief is the kind of informational state that is available to connect to other informational states and, in doing so, "inform an extremely wide range of behavior" (42). Here Kornblith clearly has in mind the functionalist idea that we define beliefs by reference to their causal role in relation to inputs, outputs, and other mental states. We need to posit the existence of such functional states, he thinks, because we cannot understand the full complexity of animal behavior in the way we understand the actions of plants, strictly by reference to their biochemistry. He invokes Fodor's (1974) discussion of the explanatory value of higher-level properties of the special sciences. Just as *camshafts* may not reduce to a single physical type but explain automotive "behavior," the functional properties of *beliefs* may be multiply realizable in many media while still providing explanations of animal behavior. And Kornblith thinks we need to appeal to this higher level property of beliefs in order to explain complex animal (especially human) behavior.

Of course this suggests that "belief" is *not* a natural kind in the sense of a natural *physical* kind. What this means for the status of knowledge as a natural kind depends on whether it is possible for a natural kind to supervene on properties that are not themselves natural kinds. Perhaps there *is* a natural kind of *animal* belief, or *human* belief, however, and Kornblith has a way out. In any case, whether or not we can really regard belief or knowledge as natural kinds, Kornblith is still appealing to a broadly functionalist naturalization strategy. In fact his particular strategy seems particularly amenable to our third naturalization proposal below (conceptually indifferent scientific naturalism). Its view of content (as an internal state that is available for manipulation to deal with the environment) is strikingly similar to some views of content (especially Cummins (1989) and Waskan (2006)) that we will examine below. Since Kornblith does not elaborate on his view beyond a few pages, we will need to look to the philosophers of mind discussed below to do the job for him.

This seems to be the case, in general. None of the major optimistic naturalistic epistemologists offer substantive account of their own of how belief is to be naturalized. Instead they rely on the proposals of others. Given the division of philosophic labor between epistemologists and philosophers of mind, this is understandable. However, as we shall now see, the philosophers of mind who endeavor to naturalize belief almost invariably have different purposes than those of the naturalized epistemologists.

Belief naturalization proposals

In the remainder of this chapter, I will outline three separate proposals for naturalizing belief, and argue that none achieves the task in the manner needed by the naturalized epistemologist. I borrow the taxonomy of naturalization proposals from Michael Tye (1992), who divides them into "analytic naturalism," "conceptually regulated scientific naturalism," and "conceptually indifferent scientific naturalism." In this section of the paper, I will draw on Tye's taxonomy to argue that none of these naturalizations proposals are satisfactory for the purposes of the naturalized epistemologist. Tye himself comes to the same conclusion, but I will produce my own reasons for thinking it, while at the same time updating and enriching the description of which more recent proposals fall under these categories.

The first category maps neatly onto the proposal of the same name for *knowledge* naturalization mentioned in the first chapter. The second and (part of the) third categories fall under what I've called "two factor semantical naturalism" about knowledge in the first chapter. So each of these proposals for how to naturalize *belief* derives from a theory of reference applied to the *concept* "belief." In the present taxonomy, what I called "supervenience" naturalism in the first chapter is subsumed under conceptually regulated scientific naturalism. (As we discovered in chapters one and two, supervenience offers very little when taken by itself, and must be given conceptual guidance.)

Interestingly, not every advocate of a particular knowledge-naturalization proposal also advocates the parallel belief-naturalization proposal. A case in point is Alvin Goldman, who unabashedly defends the purely analytic approach to knowledge, but gestures toward conceptually-regulated naturalism in his discussion of belief. It is not entirely clear what explains this in Goldman's case, but it probably expresses the fact that contemporary adherents to analytic naturalism are hard to come by, for reasons we are about to discuss. It is probably just too obvious that analytic "naturalism" is too analytic for naturalists to stomach. The second and third proposals are more popular, in proportion to the extent to which they move away from traditional conceptual analysis and towards unfettered scientific investigation.

Analytic naturalism

Tye (1992, 424) describes analytic naturalism as "the thesis that our psychological concepts have necessary and sufficient conditions for their application—conditions that may be elicited by *a priori* examination." The rationale for considering it a kind of naturalism is presumably that the proposed *analysans* would involve only concepts that are themselves naturalistically acceptable, referring either to behavior or physical stimuli or some other scientifically respectable properties.

A typical example of a traditional form of analytical naturalism is Ryle's (1949) analytical or "logical" behaviorism. According to this view, to speak of mental states of any kind (whether intentional or phenomenal) is just to talk about dispositions to engage in particular types of behavior, given particular stimuli. "Gilbert believes it will rain" just means something like "Gilbert is disposed to bring an umbrella with him." Of course there are numerous well-recognized flaws in this analysis. Gilbert may not be disposed to bring the umbrella if he wants to get wet, even if he believes it will rain. Specifying this condition of course requires reference to further mental states, which defeats the attempt to define particular mental-state attributions in neat, purely behavioristic ways.

Problems like this lent credence to the push for analytic *functionalism*, which followed the lead of analytic behaviorism by defining mental states in terms of the output (e.g., behavior) and input (e.g., stimulation) of systems, but included under each the possibility of *other mental states* (defined the same way). So, for example, to speak of Gilbert's belief that it will rain, we must refer not only to his disposition to bring an umbrella, but also his desire to stay dry.

A rigorous method of specifying functional definitions of mental states was presented by David Lewis (1972), who argued that *all* theoretical terms are defined by their causal role. A detective, for example, might define suspects X, Y and Z in terms of their hypothesized roles in a murder plot, i.e., by their individual interactions with the victim and/or their interactions with each other, where these interactions are defined in pre-theoretical or "observational" terms. When the detective asserts that a theory involving X, Y and Z is true (using a "Ramsey sentence"), he offers an implicit definition of the terms "X," "Y" and "Z" (which can be formalized with a conditional "Carnap sentence"). If the story ends up being false, the terms fail to refer (although they still have a meaning in virtue of picking out "alternative possible worlds") (252).

Lewis argues that the same style of definition can be offered for mentalistic terms. Lewis tells us to collect "all the platitudes" we know of describing the causal role of mental states in terms of other mental states, stimuli, and behavioral output (256). We then formulate our theory of the mental using Ramsey sentences, e.g. this rough functional definition of a belief that it is raining (here I reinterpret some of Lewis' formalizations in terms of the example of the functional definition of belief I have already been discussing):

∃x∃y(x is caused by rain & x is caused by the desire-for-dryness y & x causes umbrella-brining) or

$\exists x \exists y$ (Rx & Dxy & Ux)

for short. This Ramsey sentence is then turned into a meaning postulate for mental state terms using a Carnap sentence:

$$\exists x \exists y\ (Rx\ \&\ Dxy\ \&\ Ux) \supset (Ra\ \&\ Dab\ \&\ Ua)$$

Where “a” and “b” name the mental states of belief and desire. Strictly speaking, this must be paired with another conditional describing what happens if nothing fits the description,

$$\sim\exists x \exists y\ (Rx\ \&\ Dxy\ \&\ Ux) \supset a\&b = *$$

where this means that “a” and “b” fail to refer. Pairing these conditionals gives us the analytic truth that either these mental states do not exist *or* our platitudes are true (“most of” our platitudes) (257). Furthermore, we can use this analysis to *reduce* the mental to the naturalistic if we identify x and y with some other independently (naturalistically) specified entities, call them p and q, which realize the truth of the mental platitudes.

The view of reference Lewis relies on here is strongly descriptivist. As Stich has noted, it is also the view of reference that underpins the case for eliminativism offered by Churchland (Stich 1996, 29–34). Folk psychology, on his view, offers an implicit functional definition of mentalistic terms, and if folk psychology’s basic tenets prove to be false or lack predictive/explanatory power, then these terms fail to refer. This is a possibility Lewis considers: what he takes to be an analytic truth is not that folk psychology is true, but that if it is true, then “belief” refers (and if not, it does not). So one reason that analytic naturalism is not conducive to naturalizing belief is simply the possibility that the debate over the power of folk psychology might have an unfavorable outcome, and the eliminativist could win too easily. Different, non-descriptivist accounts of reference (and their corresponding naturalization projects) would permit folk psychology to be false or unreliable, but still permit folk psychological terms to refer.

Going into the debate about the power of folk psychology would be too much of a digression. What I would like to do instead is to say more about the methodology of this entire naturalization proposal. One point is that a descriptivist theory of reference is not our only option. We may instead opt for an externalist or causal theory of reference, following the Kripke-Putnam thought experiments, or a hybridized causal-descriptive theory (a two-factor theory). There is then the question of how we might naturalistically decide in favor of one theory of reference versus another. Drawing on problems similar to those noted by Quine's inscrutability of reference thesis (1969c), Stich (1996) has argued there is no naturalistic way to decide the matter. We shall return to this question later under the heading of the third proposal. For now, we shall work within this descriptivist framework and concern ourselves with whether the analytic method it relies on is consistent with naturalism.

The more immediate concern with this style of naturalization proposal is whether any proposal appealing to *analytic truth* is naturalistically acceptable, given the long-standing naturalistic animus—following Quine (1953b)—against that concept. Stich (1996, 79–80), in particular, doubts that Lewis' dependence on analyticity can be squared with a fully naturalistic outlook. This raises a question: should contemporary "optimistic" naturalists follow Quine's critique, or find some way to accommodate themselves to analyticity? In a critique of Stich, Tim Crane (1998) says that Stich moves too quickly. He notes that there are contemporary views of analyticity, in particular Boghossian's (1996), which needn't be committed to the eccentricities targeted by Quine. We should briefly examine Boghossian's theory to see if it will do the work the naturalist needs.

Boghossian distinguishes *metaphysical* from *epistemological* analyticity. *Metaphysical* analyticity is the kind sought by the logical positivists: a statement is analytic in this sense provided that it "owes its truth value completely to its meaning, and not at all to 'the facts'" (1996, 363). *Epistemological* analyticity—which Boghossian supports—concerns not the source of truth but the source of justification: a statement is analytic in this second sense provided "grasp of its meaning alone suffices for justified belief in its truth." Boghossian rejects the metaphysical concept of analyticity on the grounds the *mere* fact that a sentence S means that P could never *make* S true. Even simple identities like "Copper is copper" are true, in part, in virtue of the general fact about the world that everything is self-identical. It would be equally absurd to claim that prior to our meaning something by the sentence "Either snow is white or it isn't," it wasn't *true* that either snow was white or it wasn't (364–5). But Boghossian thinks that the metaphysical concept of analyticity can be safely rejected without rejecting the epistemological concept.

Boghossian characterizes epistemological analyticity in terms that are strikingly similar to Lewis' view of the meaning of theoretical terms. He adopts a "conceptual role semantics," according to which some expressions "mean what they do by virtue of figuring in certain inferences and sentences" (382). This is more expansive than Lewis' view, because it characterizes the meaning not only of individual terms via the sentences in which they are used, but also entire sentences via the inferences in which they figure. Boghossian thinks that theory of meaning is unavoidable for expressions such as logical constants like "not," "and," and "or": by themselves they have no distinctive "flash" meaning. As Frege emphasized, it is only in the context of entire sentences that they have any meaning at all. This conception of meaning, Boghossian says, points directly to the desired kind of epistemological analyticity: because, for example, the meaning of logical terms derives from their use in sentences and inferences that we take to be true, *if* those terms mean what they do, then those sentences/inferences have to be true/valid. Knowing, then, that the terms *do* mean what they mean, we then acquire *a priori* justification for believing in the truth/validity of the relevant sentences/inferences (and likewise in the truth or validity of any other sentences/inferences that determine the meaning of any other expressions we know).

Of course any given term or expression can be used in *many* different sentences or inferences. Boghossian realizes that the Quinean objection to conceptual role semantics is that the particular sentences or inferences that constitute the meaning cannot be isolated, but he convincingly argues that unless we assume some of them constitute the meaning, the very notion of meaning itself must be indeterminate. He notes that while most philosophers seem to agree with Quine's critique of analyticity, few agree with his arguments for the indeterminacy of meaning (Quine 1960). He urges philosophers to go with their intuitions against indeterminacy, and therefore embrace the possibility of meaning-constitution, and with it, analyticity.

But that's it. Boghossian does not dissect Quine's arguments for the indeterminacy thesis. Perhaps this approach is legitimate to philosophers who place a high premium on their intuitions. There is, after all, a plausible objection that the indeterminacy thesis is a *reductio ad absurdum* of Quine's philosophy, rather than a surprising conclusion derived from mundane premises. But to Quine, at least, the premises from which he derives his indeterminacy thesis are central to his philosophic naturalism. Meanings, or propositional objects, are intensional, and lack the precise identity conditions naturalists would like. Indeterminacy also follows from meaning holism, which follows on one level from his confirmation holism (the Quine-Duhem thesis) (Quine 1969a, 80–1), and on another from his inscrutability thesis (Quine 1970, 182). In chapter 5, I will examine his arguments for indeterminacy in some detail, and suggest that both levels of argumentation are rooted in the hypothetico-deductive view of confirmation, which has long been cherished by naturalists. If naturalists want to reject Quine's indeterminacy thesis, they will need to reject these cherished naturalistic views.

But of course, if naturalists do this, we may wonder why they would need to be naturalized *epistemologists* anymore. As we learned in chapter two, *Quine's* reasons for wanting to naturalize epistemology are inextricably connected to his indeterminacy thesis. More broadly, Quine's reasons are connected to his rejection of the possibility of *a priori* epistemology. If, however, Boghossian's view gives us a conception of analyticity that permits us a source of *a priori* justification, it is less obvious why we need to naturalize epistemology. With *a priori* justification, we ought to be able to analyze our concept of knowledge and seek first principles of its justification. So even if Boghossian's argument vindicates analyticity—and it is not clear that, to the serious naturalist, it does—it then runs the risk of proving too much. Analyticity in the service of naturalizing belief runs the risk of obviating the very need to naturalize knowledge.

As a sidebar, it is interesting to note that there have been a few attempts to naturalize the *a priori* itself, that is, to show that some natural system apart from the senses could be responsible for the justification of some of our beliefs. Georges Rey (1988) argues for a view like this by supposing that we might have a subsystem in our brain capable of "grinding out the theorems of first-order logic," one that causes its possessor to be able to believe truths of logic (33). This would count as knowledge insofar as truths of logic are true if anything is, and *a priori* because beliefs in these truths would be accepted independently of sensory experience. Rey thinks that the model of justification here is simply that of the reliabilist: beliefs caused in this way would simply be "absolutely reliable," i.e. "the result of a process that reliably issues in true beliefs in *all possible circumstances"* (34). In answer to opposition from critics, Rey clarifies that beliefs caused in this way would be more than accidentally true. He claims that if logical truths can be specified through Tarskian recursion, using logical operators and referential devices, as true-in-a-language, then the logical truths are "those sentences that are true by virtue of the pattern of operators alone, independently of the assignment of the referential devices, i.e., they are true 'merely by virtue of their logico-syntactic form'" (35). The logical-synactic properties of the subsystem that cause belief in the truths of logic, then, may be the very properties that also make them true.

In criticizing this view, I will leave aside, for the moment, the fact that it presupposes the very notion we are attempting to naturalize: the notion of belief (though this is a major problem). I want, instead, to make methodological points. The core of this proposal is, of course, to account for the *a prioricity* of logical truth—not yet for a wider concept of truth in virtue of meaning. But an immediate worry is that even if truths of logic are true at least in part because of their form, Rey's contention that they are true by virtue of this form *alone*, and not in virtue of anything in the world, is quite tendentious. It seems to involve all of the problems of metaphysical analyticity that Boghossian rejects. But even if this did provide an acceptable account of the *a prioricity* of logical truth, it is difficult to see how this account would serve the naturalist's purpose by clarifying the methodology of naturalization. The analytic naturalist, seeking to naturalize belief, needs more than an account of logical truth. Rey does offer to extend his account to include truths in virtue of meaning, by suggesting that we might also have beliefs that result from the application of our logical subsystem to certain meaning rules. Meaning rules, on this conception, would be "slots" in the "file" that functions as a concept, the slots that constitute the concept's identity by specifying rules for determining an extension (37). Now Rey acknowledges that it is very difficult to specify just which "slots" constitute the concept's identity. The problem here, presumably, is the same problem as Boghossian faces: which sentences or inferences constitute an expression's meaning, as opposed to other truths in which it figures? Rey says he is not concerned with specifying the meaning rules, just with saying that *if* we can specify them, then there is an available notion of naturalistic *a priori*. Perhaps, but my criticism of Boghossian applies here just as well. The rejection of any principled account of the meaning-specifying sentences or inferences is at the core of Quine's account of naturalized epistemology. Finding such an account is where the real work is needed, and if we had one, we may very well not need naturalized epistemology in the first place. The record of attempts to naturalize the *a priori* is not encouraging here. Others who have attempted the same strategy, such as Kitcher (2000), have also done so by characterizing the *a priori* via some kind of reliabilistic warrant, have concluded that even if there is a coherent naturalistic concept of the *a priori*, it's unlikely that there *is* any *a priori* knowledge of interest to speak of. Kitcher, in particular, argues that even mathematics may not be *a priori* under this concept of the *a priori*.

In the preceding discussion of both Boghossian and Rey, I have relied heavily on the idea that it is difficult to provide a naturalistic account of the meaning-constituting inferences or sentences of a particular term, particularly in light of Quine's indeterminacy thesis about meaning. I've said that this does not mean that Quine is right, or that there is no workable account (naturalistic or otherwise) of meaning. An interesting case is the view of Michael Devitt (1996), a serious naturalist who nonetheless espouses a "semantic localism" about meaning (as opposed to Quine's semantic holism, which leads to his indeterminacy thesis). Devitt thinks meaning can be understood naturalistically via theory of reference, and that different theories of reference can help explain different kinds of semantic behavior (descriptivism works for some concepts, but he thinks terms of descriptions must ultimately acquire reference through causal connections to the world (160–1)). So perhaps a theory of meaning can be naturalized by a theory of reference. What is important about this theory is that even if it does provide a naturalistic account of meaning, Devitt insists that it lends no quarter to a theory of analyticity. Just because meanings are real does not mean we can know them through *a priori* conceptual analysis (which makes sense if some meanings are constituted by causal connections) (18-38). So this naturalistic theory lends no comfort to either metaphysical or epistemological accounts of analyticity.

This leaves open the question of whether naturalists like Devitt could still use an acceptable theory of reference to determine, *a posteriori*, whether or not there are theoretically important properties picked out by "belief." Perhaps a functionalist-descriptivist theory can be adopted without its baggage about analytic truth. Or perhaps another theory of reference entirely will do the job. We will indirectly explore some of these other theories while looking at subsequent naturalization proposals.[17] In the meantime, we should ask: if there is some account of meaning or reference that could be exploited without reliance on analytic truth, will it help the analytic naturalist? Do theorists have *a priori* access to the *meaning* of their concepts, even if they don't have *a priori* access to the truth? Devitt has suggested that they do not, and some important evidence seems to support him.

---

[17] But as we have already noted, Quine and Stich raise pressing problems about naturalizing theories of reference. Quine's thesis of the inscrutability of reference, in particular, holds that numerous incompatible reference schemes can explain the same behavior. Devitt, in particular, does little to address Quine's in-principle critiques of meaning and reference. He quickly dismisses Quine's argument from confirmation holism, on the grounds that it depends on a kind of verificationism that he takes most philosophers to find unacceptable. This is a mistake, however, because a closer examination of Quine's corpus suggests that his "verificationism" is not the crude sort so easy to critique (Raatikainen, 2003). He also offers no independent critique of Quine's indeterminacy of translation argument, which in my view is a reformulated version of the argument from confirmation holism (see "Quine's acquiescence in skepticism"). In any case, in our discussion of conceptually indifferent scientific naturalism, we will look to see if any available accounts of reference do the job the naturalized epistemologist needs, in particular the job of accounting for the reference of advanced scientific theory.

There is a substantial body of research in cognitive psychology, usually embraced by naturalists, that is widely thought to show that we do not have access to any necessary and sufficient conditions encoding the meaning of our concepts. This tradition of research is usually thought of as beginning with the work of Eleanor Rosch, who was inspired by Wittgenstein's "family resemblance" view of concepts. Rosch uncovered "typicality effects" in subjects' application of various concepts, evidence often cited as establishing that concepts are encoded by "prototypes", long lists of properties, *most* of which must be satisfied in an instance for a concept to apply to it, rather than an exhaustive list of necessary and jointly sufficient properties. So, for example, even if it is thought that one essential characteristic of being a bird is that an organism be capable of flying, an ostrich may still count as a bird if it has *enough* of the other prototypical features of birds. Analysts could allow that our definitions may stand in need of revision, naturalists say this means that we can never predict, *a priori*, how new discoveries might necessitate new methods of categorization. This is a point that seems to stand even if the full-fledged prototype *theory* of concepts does not stand: conceptual analysis simply does not account for the propotypical aspects of our concepts. Stich (1988; 1992), Tye (1992) and Ramsey (1992) all invoke these findings from empirical cognitive psychology to dismiss the possibility of analytic naturalism in regards to concepts of the mental.

Frank Jackson (1998) offers a response to naturalists who invoke this psychological research about the difficulty of *a priori* access to necessary and sufficient conditions. He says that even if *we* cannot list all of the necessary and sufficient conditions of a concept's application, this certainly does not mean that there isn't anything that it is to *be* whatever the concept refers to. There may still be some infinitely long disjunction of properties that determines what it is to be some phenomenon, for example grooming behavior. What the conceptual analyst has to do, says Jackson, is simply to do "enough by way of conceptual analysis to make it plausible that the purely physical account of our world makes true the grooming-behavior account of our world" (62). What exactly it is to do "enough" is of course an interesting question. He mentions Lewis' functional analysis of the mental as an example, though concedes that it does not complete the naturalization project. A second stage is needed, in which the physical realizers of these functions are identified. This proposal is, in fact, at the heart of the second kind of naturalization proposal which we are about to examine.

It is just as well that we should move on to examine the next proposal, because this first (analytic naturalist) proposal is not very popular among naturalists. We have already noted that Goldman, although an advocate of analytic naturalism about knowledge, seems to favor a different proposal for belief. Even Lewis (1995), whose view of the meaning of theoretical terms is the most conducive to analytic naturalism, admits more recently that it offers only a "recipe" for analysis. Later on he seems to side more with Jackson on how that recipe is to be completed.

Conceptually-regulated scientific naturalism

Tye (1992) has described "analytic naturalism" as searching for *a priori* necessary conditions of the mental. In the preceding section, we have found that this proposal faces severe methodological problems as a naturalistic approach to belief and the mental, more generally. Not only does it seem to be difficult to naturalize *meaning* in the way a functional analysis of the meaning of "belief" would require, but it seems equally difficult to naturalize the kind of *a priori* access needed to exploit knowledge of that meaning for philosophic purposes. Because of considerations like this, Tye describes a second category of naturalization proposals that is intended to overcome the problems of the first. He calls this second category "conceptually regulated scientific naturalism," which he describes as follows:

> Scientific investigation, together with philosophical reflection regulated by our pretheoretical conception of mental states, is needed to come to a full understanding of their essences. (424)
>
> According to [this view], mental state types have non-mental essences. The task of the philosopher of mind is to specify what *sorts* of [non-mental] essences these are and correlatively to say which sciences will discover them, the primary constraint on any acceptable proposal being that it must be compatible with our ordinary, pretheoretical views about where the boundaries of the mental lie. (426)

Immediately we can think of some early naturalization proposals that might have fit this description. J.J.C. Smart's (1959) identity theory is one example. Smart thought that scientific investigation had revealed that the referents of our concept "pain" could be (contingently) identified with the stimulation of C-fibers. Likewise a Nagel-style reduction of the mental to the physical, which would connect primitively understood mental predicates to physical predicates through "bridge laws," would probably also count as this variety of naturalism. Both of these proposals would qualify as "type-type" physicalism, according to which a pre-theoretical mental type (e.g., pain) might be identified with or reduced to a physical type (e.g., C-fiber stimulation).

Of course towards the end of the 20th century, type-type physicalism lost much of its popularity. "Multiple realizability" arguments, in particular, suggesting that the mental could never be identified with or reduced to a single physical type, because there could be beings in other possible worlds—or even unknown beings in this world—which realized mental properties without having the same neurophysiological basis as ours (Putnam 1975; Fodor 1974). The question then became how to naturalize the mental in lieu of finding a distinctive physical property containing within it the key to all mentality. Of course the multiple realizability problem contained the seeds of its own solution: there must be some way in which we would be able to *recognize* these multiply-realized properties as mental, and the usual candidate is a *functional* criterion. So this much the newer approach shares with the analytic naturalist. But explaining how we identify properties as mental is not enough to *naturalize* them. More is needed, as Jackson observed at the end of the last section, to show how these mental tokens are *realized* in the physical. So perhaps mental types cannot be identified with or reduced to physical types, but they may *supervene* on the physical, as realized in functional types.

Tye gives a few somewhat unconvincing examples of naturalists who seem to fit the mold of his "conceptually regulated scientific naturalism. But clearly Jackson's idea of *beginning* with conceptual analysis (of the functionalist variety), and proceeding to find the physical realizers (of whatever type) of those functional types is a paradigm. Writing in 1992, Tye might not have been able to predict the rise of an entire school of philosophic methodology, later in the 1990s, that would supply a semantics just for this view of naturalization. I am speaking of the "two-dimensionalist" semantics of Frank Jackson (1998) and David Chalmers (1996). Chalmers, in particular, applied the semantics to questions in the philosophy of mind. His approach seems to have been endorsed more recently by Jaegwon Kim (2005).

Each of these thinkers makes special use of the concept of "supervenience." Both Chalmers and Kim argue that while supervenience does not *reduce* the mental to the physical, it does offer a "reductive explanation" of the mental (Chalmers 1996, 43; Kim 2005, 93–120). According to Chalmers, a phenomenon such as belief is reductively explained by lower-level natural properties when it *logically supervenes* on those properties (1996, 47–8). Logical supervenience, in turn, is understood as follows: "[A]-properties supervene *logically* on [B]-properties if no two *logically possible* situations are identical with respect to their [B]-properties but distinct with respect to their [A]-properties" (35).[18] Less formally, B properties *determine* A properties.

Kim (2005, 101–2) offers a useful schematization of the steps taken in a process of reductive explanation:

> STEP 1 [FUNCTIONALIZATION OF THE TARGET PROPERTY]
> Property M to be reduced is given a *functional definition* of the following form:
>
> Having M = def. having some property or other P (in the reduction base domain) such that P performs causal task C.
>
> STEP 2 [IDENTIFICATION OF THE REALIZERS OF M]
> Find the properties (or mechanisms) in the reduction base that perform the causal task C.
>
> STEP 3 [DEVELOPING AN EXPLANATORY THEORY]
> Construct a theory that explains how the realizers of M perform task C.

I would now like to describe these steps in more detail, drawing in particular on Chalmers and his semantics.

[18] I have inverted Chalmers' "A" and "B" here to bring the formulation in line with earlier discussions of supervenience in this dissertation.

Chalmers says that the first step, “functional analysis,” requires only a “rough and ready” analysis, and that it is common—and necessary—to begin a variety of reductive explanations in science in this manner. He gives the example of “reproduction”: without an analysis of “reproduction” as some kind of “ability of an organism to produce another organism in a certain sort of way,” science could never ascend from descriptions of relationships between complex entities and explain how these entities *reproduce* (1996, 43–4).

Now one might claim at this point that the same objections raised against the conceptual analysis of the analytic naturalist apply in equal measure to the first stage of this conceptually regulated reductive explanation. Problems included an unaccounted-for concept of meaning, uncritical reliance on a descriptivist theory of reference, and difficulty of *a priori* access to the necessary and sufficient conditions of concept application. But Chalmers has a more sophisticated view of conceptual analysis which offers answers to each of these objections. The sophistication comes from his view of semantics, which supports each of his responses. Notice that Chalmers makes liberal use of the concept of “logical possibility.” In the earlier chapter on Kim, we explored how any notion of supervenience relies on *some* conception of possibility or necessity, and found that Kim’s attempt to couch it in terms of nomological necessity was largely unsatisfactory. Chalmers’ invocation of strict logical possibility bypasses that problem, but as a result he owes us an account of logical possibility and necessity. The account he will present is the same that fills out his distinctive view of conceptual analysis: it is a traditional possible worlds account (57).

When Chalmers describes the condition of supervenience as that in which no two logically possible situations are identical with respect to B-properties but distinct with respect to A-properties, he means there is no possible world with the same B-properties as ours but with different A properties (70). His account of the analysis of the *concept* of an “A” also invokes possible worlds. Chalmers divides the meaning of a concept into two “dimensions” (hence this is a “two-dimensionalist” semantics). The first dimension, which he calls “primary intension,” is “a function from worlds to extensions reflecting the way actual-world reference is fixed" (57). He makes use of Putnam's example of the concept "water" to illustrate this function. Our primary intension of water picks out the extension the "dominant clear, drinkable liquid in oceans and lakes" in each possible world: we say that water is H_20 in our world, that it is XYZ in another possible world, and even that water is *both* H20 and XYZ in a possible world if one occupies our lakes and the other occupies our oceans (57–8). This, then, is how Chalmers' primary account of the notion of meaning. Its only substantive philosophic assumption is the notion of possible worlds as primitives (66).

We can now appreciate Chalmers' responses to the second concern about analytic naturalism, regarding its descriptivism. Although he uses the example of "dominant clear, drinkable liquid in oceans and lakes" to illustrate how primary intension works, it is not the description so much as our dispositions to call things water that matter. "It is the function itself, rather than any summarizing description, that is truly central," Chalmers tells us (59). He says this is compatible with a causal theory of reference if our reflections on our dispositions leads us to think that reference is secured by a causal connection. For example, reflecting on our disposition to call watery stuff "H_20" on Earth, but "XYZ" on Twin Earth might lead us to formulate a causal theory of reference, given that we are in causal contact with the former substance on Earth, but not on Twin Earth. This is why I say that Chalmers' view is possibly viewed as a "two-factor" theory of reference. Usually causal theories of reference need supplementation with descriptive aspects, so if Chalmers permits a causal element, it is likely he would admit that both causal and descriptive factors function to achieve reference.

Finally, the contrast between primary and *secondary* intension allows him to account for the particular results of the Twin Earth thought experiments that are thought to undermine *a priori* access to important necessities. Kripke (1972) argues that sentences like "Water is H_20" are necessary but *a posteriori*: given the empirical discovery that water is H_20 in the actual world, it is H_20 in all possible worlds. Chalmers' response is that primary intension, being a function from worlds to extensions, is held in essentially conditional form: *if* watery stuff in our world is H_20, then it is water; if watery stuff in another world is XYZ, then it is water, etc. This much, he says, is *a priori*, as it is determined merely by reflection on our speech dispositions. What Kripke is correct about is the *secondary* intension of water, which depends on the primary intension. Using the primary intension, when we learn that watery stuff is H_20 in the actual world, we fix the reference of "water" in the actual world, but then "rigidify" it and grasp that water is H_20 in all counterfactual possible worlds. Because if water is *this stuff* in our actual world, and this stuff is H_20, nothing that is not H_20 in other possible worlds can count as water, even if it is watery stuff. The trick, according to Chalmers, is that this *a posteriori* necessity depends on the *a priori* analysis of the primary intension. So while "Water is H_20" is not *a priori*, "Water is watery stuff" is (62). This is all Chalmers needs for his reductive explanation of the mental, because he thinks that the functionalist definition of belief is comparable to "Water is watery stuff" (79).

Understanding primary intension as a function from worlds to extensions also enables Chalmers to answer concerns raised by naturalist psychology about definitions and our *a priori* access to them. Chalmers acknowledges, of course, that crisp definitions in terms of necessary and sufficient conditions are not always available. But he argues that verbal definitions in terms of necessary and sufficient conditions are only often useful summaries of the meaning of our concepts, not the meanings themselves (78). The kind of meaning relevant to reductive explanation is primary intension, which is not a description but a function, expressed by our speech dispositions. The prototypicality effects noted by psychologists reveal what our speech dispositions are; as such, they reveal something about our primary intensions. For this reason they are no problem for conceptual analysis.

Having presented Chalmers' answers to objections to conceptual analysis, we have completed our description of Kim's first stage of reductive explanation, i.e., functionalization of the target property. Surely for theorists attempting to formulate a reductive explanation of belief, some verbal statement of a functional definition of "belief" is necessary. But Chalmers' point is that whatever the limitations to the process of definition, they are not significant, given that the theorist has more immediate access to his primary intension. Having access to that primary intension, he can now move to the second and third steps of reductive explanation: identification of the realizers of the functional property.

Identity and reductionist theories have stumbled at this second step. Chalmers alleges that reductive explanation does not stumble, for two reasons. First, reductive explanation does not need to explain by reference to *types*: we need only offer reductive explanations of *tokens* of higher-level phenomena like belief (43). Supervenience provides the apparatus to offer this kind of explanation: a *particular* higher-level functional property can be said to supervene on a lower-level property if, already being in possession of the concept of the higher-level property, we can infer it from knowledge of the lower level property (76), or if we simply cannot conceive of a world with the lower level property *without* the higher-level property (73). Second, Chalmers thinks that the lower-level property need not be physical, strictly speaking. In our second step, we can descend to the level of neurophysiology, positing neurophysiological states as the realizer of functional properties. If we then explain how human neurophysiological states perform the functions in question, we will have completed our reductive explanation (without having to commit to the idea that all possible beliefs are realized in the same way) (46). But we also do not *need* to descend as far as the neurophysiological: he suggests that cognitive science could offer more "abstract" models of mechanisms giving "how-possible" explanations in terms of the known causal organization of organisms (46).

Chalmers thinks that such explanations are in principle available for psychological concepts like learning and belief—though notoriously, he thinks they are not possible for phenomenal concepts (because we can imagine zombies). As a result, Chalmers thinks that intentional psychological concepts like belief are "straightforwardly logically supervenient on the physical": whatever lower-level properties we identify as the realizers of belief, we cannot imagine beliefs differing where these realizers do not (82–3). Of course Chalmers thinks that because we can imagine conscious properties (phenomenal qualia) differing without physical differences (because of zombies), these do not supervene. And he recognizes the possibility that intentional content may itself depend on phenomenal qualia, in which case intentional properties themselves might not supervene. But he thinks there is at least a third-person version of intentionality available that does not depend in this way. This version of intentionality, therefore, he takes to be supervenient. Supervenience, according to Chalmers, is the guide to judging the place of a phenomenon in the natural world (32). In this sense, we can say that Chalmers' theory of reductive explanation offers a *naturalization* of belief, or at least a proposal for how a naturalization might be achieved (and thus a plausibility argument for how it *will* be achieved). Insofar as this style of naturalization begins with *a priori* analysis of a primary intension, and ends with an identification of a non-mental essence, it looks like a good example of what Tye calls conceptually-regulated scientific naturalism.

Having outlined the most compelling proposal for a conceptually-regulated scientific naturalism (Chalmers'), we are now in a position to question its naturalistic credentials. First we need to question whether the concept of meaning (primary intension) exploited by this view is naturalistically respectable. Then we need to determine whether primary intension, even if naturalistically respectable, can be readily accessed in the *a priori* manner that Chalmers insists. Finally, we will look at likely candidates for the supervenience base, and whether or not they themselves can be naturalized.

As we have mentioned, Chalmers' reliance on both the notion of supervenience and the notion of primary intension depends on claims about logical possibilities and necessities, which claims are interpreted by reference to talk of logically possible worlds. In our earlier chapter on Kim's critique of Quine, we mentioned his reliance on the concept of epistemic supervenience and the naturalistic problems it faced. Having dismissed possible-worlds based accounts of possibility and necessity as hopelessly non-naturalistic, we noted how Kim attempted to characterize supervenience in terms of a *nomological* notion of possibility and necessity, but this proved cumbersome and implausible. Why, then, were we so convinced that possible worlds semantics was incompatible with naturalism? Much of the reason is the presumption that talk of necessity and possibility generally appears to be irreducibly intensional (with an "s"). Quine noted as early as "Two Dogmas of Empiricism" that terms within the scope of modal operators fail to be intersubstitutable *salve veritate*, and lack scientifically respectable extensional identity conditions (Quine 1953b; Quine 2004a). Since Quine is a paradigm naturalist, there is a certain presumption that skepticism about intensionality is a hallmark of the naturalist outlook. Talk of possible worlds would seem to feature the same problems, and probably more—given that it also seemingly relies on *a priori* access to these worlds, which naturalists are also likely to doubt (Brandom 2001, 598; Moreland 1998).

But Chalmers presents his reliance on possible worlds in a manner that some naturalists might find compelling. While he says that the notion of a logically possible world is to be treated as "something of a primitive," he also says that we should treat them "as a tool, in the same way one takes mathematics for granted" (Chalmers, 66). This is a provocative answer to the Quinean critique of modality, because Quine himself was quite content to treat mathematics in just the manner Chalmers suggests, as naturalistically acceptable if only because of its pragmatic indispensability for doing science. Now Robert Brandom (2001, 599) responds specifically to this practical indispensability argument. He says that it may very well be that actual scientific practice relies on modal notions, and for this reason, they must be seen as pragmatically indispensable. But he denies that this is sufficient reason to generate naturalistic respectability for use of modal terminology in *semantics*. Philosophical semantics needs to be more self-conscious and critical, in order to adjudicate the legitimacy of modal notions.

I presume that Brandom's reason for thinking this is that if we can be instrumentalists about the modal, we might as well be instrumentalists about "belief" or even about "knowledge," or any other philosophically controversial concepts. The point is that modal notions are at least as philosophically controversial as these mentalistic concepts—*and for exactly the same reasons* (both modal and mentalistic concepts are intensional). Given the equal amount of controversy, an instrumentalist about modality would need to present a special reason for which that controversial concept needed no realistic naturalization, while the others did. I take it that optimistic naturalized epistemologists, who are realists about more than just belief, do want more than instrumentalism about belief, and certainly more than instrumentalism about knowledge. So it seems to follow that naturalists will need to produce some account of modality which is, if not a modal realist account, then at least an account that provides a special reason for which modality can be treated instrumentalistically, while at the same time situating it in a scientific context in a way that is consistent with our being realist about other things.

Genuine modal realism runs up against a variety of traditional naturalist and empiricist objections tracing back not just to Quine, but to Hume. This leaves the possibility of modal fictionalism, which I briefly examined already in chapter 2. If naturalists can sketch a naturalistic account of modal fictionalism, perhaps modality itself can be "naturalized," in effect providing us with our special reason for treating it in a non-realistic manner. Indeed there are theories of modal fictionalism available in the literature (see, e.g., Rosen (1990)). According to these views, literally speaking, existence claims about possible worlds are false; there is only the actual world. But possible worlds talk is really shorthand for literally true statements about certain convenient fictions. Translations for the shorthand would look something like the following (courtesy of Nolan (2002)):

Possibly P iff according to the fiction of possible worlds, P is true at some possible world.

Necessarily P iff according to the fiction of possible worlds, P is true at all possible worlds.

Presently, however, I will argue that these fictionalist accounts do not succeed in a manner favorable to the goals of the naturalized epistemologist.

One concern noted by critics of modal fictionalism is that the translations above count crucially on understanding "according to the fiction of possible worlds…." Rosen offers several possible translations for this construction: "If PW were true, then P would be true; If we suppose PW, P follows; It would be impossible for PW to be true without P being true as well" (1990, 344). But as

Nolan (2002) observes, these translations invoke modal concepts themselves, and would render modal fictionalism a circular explanation for the literal truth of modal claims. Attempts to reduce these modalities to a primitive modality also do nothing to advance the explanatory value of this account. Modality is still modality.

Even if a version of modal fictionalism could eliminate any concerns about circularity, an even more pressing concern looms for the usability of such an account by the naturalized epistemologist. What sense is to be made of "according to the fiction"? As Nolan (2002) notes, most presentations of modal fictionalism proceed on the assumption that modal fictions operate just like ordinary fictions (say, about Sherlock Holmes). But this of course raises questions about the ontology of fiction itself. Normally we would think about fiction as being a kind of counterfactual *representation*, a portrayal of how things might be but are not. Fiction even seems to have intentional content: stories about Sherlock Holmes are *about* a detective. But by supplementing theory of supervenience of the mental on the physical with an account of the language of possible worlds, it is representational content that we are trying to explain.[19] Thus it seems that appealing to modal fictionalism to account for possible worlds talk in a naturalistically respectable manner is a non-starter for the naturalized epistemologist, even if it could serve other purposes outside of naturalizing epistemology.[20]

---

[19] Of course Chalmers is interested in explaining *doxastic* representational content, which is obviously not the same as fictional representational content. Presumably fictional representational content would not have the same world-word relations as the doxastic kind (which is why it is fictional). But it almost seems that fictional content itself presupposes doxastic content: witness the extent to which fiction is often understood as involving "suspension of disbelief."

[20] It is perhaps worth mentioning that another naturalization proposal would be disqualified if the modality of possible worlds is truly unacceptable to naturalism. Although J.J.C. Smart's identity theory was widely rejected by philosophers of mind later in the 20th century, new versions of type-type physicalism eventually arose, exploiting Kripke's idea of *a posteriori* necessity. According to Block and Stalnaker (1999), "Pain = C-fiber stimulation" and "Consciousness = pyramidal cell activity" function much like Kripke's "Water = H_20." Since these are *a posteriori* necessities, they do not require the kind of conceptual analysis described in Chalmers' account. So technically speaking they are not conceptually regulated scientific naturalism, but conceptually indifferent. However, I do not plan to address this proposal under the next section. I mention it here because, by relying on Kripkean identities, the view obviously also relies on possible worlds semantics, and should be questioned as a legitimate form of naturalism for this reason.

None of this is to say that the style of reductive explanation described by Chalmers, one that relies on supervenience and possible worlds-semantics, is unworkable *per se*. Likewise for the workability of the concepts of supervenience or possible worlds-semantics on their own. It is only to say that they do not seem to be ideal candidates for naturalization methodologies.

Although the status of Chalmers' methodology seems to disqualify conceptually-regulated naturalism on its own, perhaps there are other brands of conceptually-regulated naturalism available. There are numerous advocates of versions of functionalism that presuppose that the functional properties of belief somehow supervene on naturalistically acceptable properties. Supposing the possibility that there are other views available, I need to raise a second objection to these proposals. Even assuming they find a methodology aside from Chalmers' to determine the nature of the supervenience relation, there are still important naturalistic objections to raise about the alleged supervenience *base*.

Recall, of course, that according to Chalmers—and many agree with him—we are to search for the realizers of belief by searching for neurophysiological or other cognitive systems that have a certain functional *causal* role. A number of different philosophers propose different types of causal role that could realize these functional properties. But what is the naturalistic status of "causality"? Chalmers is aware of this problem. He notes that by his own account of supervenience, laws of nature and facts about causal connections do not supervene on physical facts (86). That is, for the typical Humean reasons, we can imagine the course of nature departing from regularities we have observed in the past. Now a popular response to the Humean problem is to explain causal concepts in terms of counterfactuals and other modal notions. While this response may be legitimate to a non-naturalist, it seems highly dubious to the naturalist, given that counterfactuals are usually then expressed in terms of possible worlds. Chalmers says that he is willing to including physical laws in his supervenience base, on the assumption that there is "something irreducible in the existence of laws and causation," but admits that this "steps over the metaphysical puzzle rather than answering it" (86). Once again, this may be a legitimate move for the non-naturalist, a naturalist who is willing to treat so many notions as irreducible may start to wonder why he couldn't just treat belief, or even *knowledge* as irreducible. This would of course obviate the project of naturalized epistemology.

At this point it might be objected that surely causal and nomological notions could be treated instrumentalistically, insofar as scientists surely rely on them constantly. Likewise, scientists surely seem to rely upon counterfactual conditionals in the very practice of setting up experiments, viz. "If this and this were set up, such and such would occur." Both of these points are true, but the question concerns their significance for the *metaphysical* status of these concepts. For these concepts to be effective in the description of the supervenience base for intentional concepts, they must of a metaphysical status that is clearer and less controversial than intentional concepts themselves, and the possibility of giving pragmatic equivalents of them does nothing to establish their metaphysical clarity. As Hume would have claimed, understanding causal and nomological concepts as reporting mere regularities is consistent with their pragmatic indispensability. Hume's own "skeptical solution" to the problem of induction—an early effort at naturalized epistemology, if ever there was one—was to say that our understanding of constant conjunction was merely a matter of "custom and habit," but not a reflection of any metaphysical relationships in the world. Yet Chalmers needs causal and nomological relations to be metaphysical if he is to exploit them as a supervenience base. Likewise, Quine himself (1994) recognized that a universally quantified truth-functional conditional could help explicate scientific experimental language, without resort to counterfactuals or other modal notions. If counterfactuals are really needed for anything (perhaps in the description of scientific laws), he also thought some nuanced version of a truth-functional conditional ("with a complex antecedent some of whose clauses are left tacit, to be divined from context and circumstances" (149)) could do the job. This pragmatic explication of certain scientific concepts does nothing to help secure the respectability of a metaphysical supervenience base, however. Whether it is realistic to fully expunge intensional concepts from scientific practice is, of course, a controversial question (see Hookway (1988)). If it is not, this may well serve as an effective critique of Quine and of naturalism. Before such a critique succeeds, however, we have to recognize the indebtedness of naturalistic philosophy of mind to these putatively non-naturalistic concepts.

The problem of the modal status of causal concepts cascades into the various theories of the *content* of belief that have been proposed by avowed naturalists. According to functionalism, "belief" is defined implicitly by reference to its causal role, which includes its input, output, and relation to other beliefs. Particularly because of widespread dissatisfaction with descriptivist theories of reference, naturalistic theories of the content of belief usually appeal to a causal theory of reference to account for the "input" end of belief's causal role. Consider, for example, Jerry Fodor's causal covariationist theory of content (1987). According to this theory, cognitive content is determined by the reliable causation of mental tokens by the properties they are about. The problem for the naturalist with Fodor's theory is not simply that it appeals to the notion of causality. The deeper problem is that to make his theory plausible, to show that not everything that causes a mental token counts as a case of successful reference, he must also account for the possibility of misrepresentation or error. And to do this, he must find a way to "idealize" the causal covariation: it counts as successfully referential only under certain circumstances. For example, suppose "mouse" is reliably tokened by mice. A subject sees a mouse and tokens "mouse." This counts as successful reference. But if a subject sees a *shrew* and mistakes it for a mouse, he still tokens "mouse." This is not successful reference, according to Fodor, because if mice didn't reliably cause "mouse" tokens, shrews wouldn't either. But as Brandom (2001, 591) rightly observes, this appeal to counterfactuals once again requires robust modal resources.[21]

The problem Fodor encounters with misrepresentation is a problem concerning the normativity of intentionality. Other putatively naturalistic theories of content have proposed dealing with the problem through other means, which can also be described as broadly functionalist. Rather

[21] For problems that Fodor's theory faces on its own terms, see Cummins (1989, 55-66).

than focusing merely on proximate causes as a source of referential "input," these theories consider the broader historical influences on an organism's representational content, particularly evolutionary influences that determine *teleological* facts about the organism. As Sober reminds us, it can be useful to put the "function" back in functionalism (Sober 1985). Prominent teleosemantic theories have included Dretske's (1986) and Millikan's (1984). In a memorable example from Dretske, we learn about marine bacteria called with internal magnets (magnetoseomes) that work like compass needles to cause the bacteria to move deeper in the water in the Northern Hemisphere, because of the direction of the Earth's magnetic field. This behavior is explained by the evolutionary advantage of seeking deeper, oxygen-free waters, which the bacteria need to survive (Dretske 1986, 26). Dretske concludes that the function of the magnetosomes is to indicate the presence of oxygen-free water. There are problems about how best to specify the function of the magnetosomes here, some that place its representational content in more distal features of the environment, some that place it in more proximal features.[22] Regardless of how the theory is formulated to specify the function, *any* specification of function requires an account of teleology which, it turns out, depends on crucially modal concepts. Recent proposals for naturalizing teleological functions (such as Wright (1976)) look like the following "etiological" account (courtesy of Neander (2004), quoting Wright (1976, 81)):

> The function of *X* is *Z* if and only if,
> 1. *Z* is a consequence (result) of *X*s being there,
> 2. *X* is there because it does (results in) *Z*.

To understand "consequence" and "because," however, philosophers exploiting this theory of teleology will resort to counterfactuals. Brandom (2001, 594) formulates the typical gloss as applied to Dretske's example: "if it *had not* been the case that the mechanism…*successfully* led to the ancestors of contemporary bacteria to less toxic waters, then they *would not* now respond as they do to magnetosomes." So the usual naturalistic objections may be raised again.

[22] Even though the teleosemantic view was formulated in part because of the problem of misrepresentation faced by the causal covariance theory, this problem of how to specify the function leads to a misrepresentation problem of its own. See Cummins (1989, 73-75).

There are naturalistic epistemologists who embrace naturalistic accounts of normativity, even when they do not follow the teleosemantics view of content. They are, instead, more interested in naturalizing normativity for the sake of understanding the normative concept of justification, especially via the concept of reliability. Often accounts of cognitive evolution are used to account for this source of normativity (Kornblith 2002, 68). For this reason, I would like to mention some more general problems for understanding teleological functions in purely naturalistic terms, problems which arise apart from concerns about modality.

Marc Bedau (1991) grants the overall appeal of etiological theories of teleogy, but argues that they contain flaws which cannot be eliminated without bringing in non-naturalistic considerations. Bedau points to the example of certain clay-based crystals that seem to fulfill all of the criteria for natural selection-based teleology that Wright addresses. These crystals are produced through as chemical processes cause molecular structures to copy themselves. When small bits are broken off, these act as "seeds" to grow again into bigger crystals. Occasionally small defects in crystals occur due to external interference, but when new crystals are created from portions containing these defects, the reproduced crystal contains the same new structure. Furthermore, some of these structures are more stable than others, meaning that some endure better than others. Yet we do not want to say that these new structures function *in order* to permit the crystal to endure longer. It seems wrong to apply biological teleology to crystals, even though they exhibit many of the superficial traits of evolution through natural selection. I believe that there is a theory of natural teleology available, developed by Harry Binswanger (1990), that adequately explains why the action of cellular respiration exhibits teleological functioning, while the growth of crystals does not. Interestingly, however, it is of no use to a naturalist interested in using teleology to naturalize the normativity of intentionality, because it presupposes this intentionality.

According to Binswanger, we first come to grasp teleological concepts by grasping our own purposive behavior. We then project purposes onto other animals (children and primitives go too far, and see purposes in plants and insentient nature), and explain their behavior in those terms. So teleology has its origins in our grasp of conscious purposes, but since only living beings are known to have these conscious purposes, it is intimately connected to the functioning of living beings. Eventually scientists find that they can formulate a concept of teleology ("goal-causation") that applies to non-conscious beings (such as plants): past instances of an action (e.g. cellular respiration) contribute to the survival of an organism, which in turn causes furtherance of the action (e.g. cellular respiration). In fact they understand this mode of teleology by analogy to purposive teleology.[23] This much of the theory resembles Wright's etiological account. But because our concept of teleology originates in the grasp of our own conscious purposes, this helps to keep even the non-conscious concept of teleology anchored to the biological. Crystals are not biological; therefore they do not count as teleological. There is a bigger story to be told here about why the concept of the teleological may be extended to the non-conscious but not to the non-living. According to Binswanger, it has to do with the explanatory power of goal-causation applied only to living beings (it explains not only superficial actions, but every aspect of their structure, down to the constant need for action at the cellular level). The main point, however, is that while this theory solves the problem of Bedau's crystals, and satisfies the judgments of biologists, it does so through an account of teleology that presupposes an understanding of an intentional concept ("purpose"), which is not available to the naturalist seeking to understand intentionality via teleology.

It looks, then, that there are serious difficulties for the naturalistic respectability of both the supervenience relation and important elements of the supervenience base, including everything from facts about causality to facts about teleology. But perhaps there is a naturalized account of modality that I do not know about. In that case, perhaps, naturalists could defend Chalmer's use of

[23] There are two aspects to the analogy. First, there is a commonsense analogy between the ontogeny of purposive action and etiological teleology: past instances of desire-satisfication also explain forward-looking desires in purposive teleology. Second, there is the history of the discovery of the theory of natural selection applied to phylogenetic teleology: Darwin himself understood natural selection in part because of an analogy to artificial selection, i.e. the purposive behavior of animal-breeders.

supervenience and primary intension. Even in that event, however, a final question to ask is: is there good reason to think we really have *a priori* access to primary intension, such that we can determine whether or not belief logically supervenes on a suitably naturalistic base?

Schroeter (2004) observes that there are a number of contemporary attempts to vindicate the methodology of conceptual analysis besides Chalmers' and Jackson's two-dimensionalist semantics, all of which allege to overcome difficulties with traditional forms of conceptual analysis. Most of these views take seriously the challenge posed by the Kripke/Putnam thought experiments, which they take to show that reference depends in some way on external factors, and hence that meaning "ain't in the head." The task for the new versions of conceptual analysis is to show that even if unpredictable externalist factors determine reference in some respect, there are other respects where this is not the case. In Schroeter's terminology, contemporary conceptual analysts concede that we do not have *a priori* access to the complete *applicability conditions* of our concepts, the complete truth about what it takes for something to fall under a concept in a given possible world, or the "semantically basic features" for a concept. But they do insist that we have *a priori* access to what Schroeter calls a concept's *reference-fixing conditions*. Rather than specifying the semantically basic features, reference-fixing conditions instead offer a generic "recipe" for determining the applicability conditions, usually via implicit metaphysical and epistemological assumptions. For example, these analysts tell us that the externalist thought experiments reveal that we have an *a priori* commitment to a *sortal* for various concepts, e.g. that water is a natural kind, one that is predominantly found in a certain state of matter having in certain locations, with a certain color, etc. Also the thought experiments reveal an *a priori* commitment to the idea that reference is to be fixed through our causal interaction with that natural kind. All of these commitments are said to be grasped *a priori* because we find that prior to empirical investigation, our strongest disposition is to call any substance "water" that meets these conditions. We may later discover that water in our world is actually H_20, and in that case we will call only those substances in counterfactual worlds that are H_20 "water," even if they do not fit meet the same sortal conditions. But this is possible only in virtue of exploiting our *a priori* grasp of the reference-fixing conditions, first in relation to the actual world. Clearly Schroeter's "reference-fixing conditions" are functioning in much the same way here as Chalmers' "primary intension."

But Schroeter argues that this new picture of the *a priori* abandons infallibility about applicability conditions for an equally difficult infallibility about reference-fixing conditions. Let me illustrate what I take to be her objection by modifying her example about water to fit something Chalmers says about its reference-fixing conditions. Although Chalmers verbally summarizes the primary intension of "water," as the "dominant clear, drinkable liquid in oceans and lakes," he sees this as consistent with our *a priori* judgment that the ice and water vapor are also made of water. Presumably this is because he thinks that the sortal for water not only includes that it is a natural kind, but a specific type of natural kind that retains its identity through state changes. But imagine that we are Empedocles, and think that water is one of the four elements. It is hard to say if thinking of water as one of these elements is the same as judging it to be a natural kind in the way we do, but it is clearly the sortal under which Empedocles classifies it. For this reason Empedocles (or some other less-sophisticated ancient Greek) might not be inclined to say that water vapor is *water*. He might think water vapor is what we find when water transforms into *air*, or perhaps some mixture of water and air. What's more, ancient Greeks uninfluenced by modern psychophysics might not be inclined to say that what counts for successful reference is a causal connection to its referents. Aristotle, for example, seemed to think that mind could not be blended with the body at all, for fear "admixture" might hinder its potential to take in the intelligible form of every possible object of thought (Aristotle 1941, III:4).

Or imagine a case in which both metaphysical and epistemological assumptions combine to render judgments about "gold" very different from ours. Imagine a philosopher like John Locke, who is explicitly skeptical about the possibility of real essences. Even if he were told a story about the atomic number of gold, he might never agree that the substance we call gold could be a vapor (of the kind we now claim to be used in certain lasers), because of his conviction that the reference of the term is fixed by a nominal essence specifying a yellow, malleable metal. If anything, it seems like assumptions about metaphysics and epistemology are even *more* variable than assumptions about gold or water. Recent work in "experimental philosophy" suggests that intuitions about reference are culturally idiosyncratic (Machery et al. 2004), which should be expected, given that not even philosophers agree on theory of reference. These intuitions seem, therefore, to be little help in specifying a useful method of *a priori* analysis. As Gary Ebbs (2003, 252) notes in a similar critique of contemporary conceptual analysis:

> The main problem with this proposal is that what we *actually* say when we find ourselves in a previously imagined situation almost always trumps our earlier speculations about what we *would* say if we *were to* find ourselves in that situation.

But Schroeter is not being entirely fair to the conceptual analysts. Perhaps they need only appeal to an *a priori* account of *justification* here. To say that these intuitions are *a priori* is not necessarily to say that they are infallible, but simply to say that they are independent of experience. Perhaps this is the view of some of the conceptual analysts. But I have two responses. First, this view may beg the question. The cultural idiosyncrasy of folk semantic intuitions suggests that they are *not* independent of experience, but learned, instead, from a predominant cultural-philosophical milieu. So there is even less reason to think that folk semantics is innate than there is to think that folk theories of gold or water are innate.

Second, even if our intuitions about reference are *a priori* in the sense of being independent of experience, this does not yet mean that they offer *a priori justification*, which is what the present group of naturalists wants—they want a conceptually-*regulated* scientific naturalism. It may be that we always need to start with some view about reference-fixing conditions to engage in any inquiry at all, but this may be only because we need to start with some view or other before we can acquire a justified view (after engaging in "reflective equilibrium"). This proposal, in fact, sounds very much like the final naturalization proposal, to which we shall now turn.

Conceptually indifferent scientific naturalism

Just because naturalists tire of looking for forms of *a priori* approaches to naturalization does not mean there is no other approach. We will consider one last approach, which Tye calls "conceptually indifferent scientific naturalism." According to this approach,

> mental states may well turn out not to have most of the properties we ordinarily attribute to them. Moreover, even if they don't turn out this way, it is certainly not metaphysically or conceptually necessary that they have such properties; and neither is it sufficient. So, any conclusions we draw from thought experiments which rest on intuitions we might have about the mental states of non-actual creatures in the light only of our ordinary, everyday conception of those states may well be in error. In matters of the mental, science, together with philosophic theorizing based on it, can be our only guides. (Tye 1992, 427)

Now Tye himself goes on to criticize this approach on the grounds that indifference towards our intuitions causes our naturalization proposals to go "out of control." "Why, on earth, should we accept a view that goes so directly against what we pre-theoretically suppose?," he asks. But Tye's criticism here is question-begging. By presuming that we need conceptual regulation to constrain our naturalizations, he assumes that conceptually-regulated naturalism is viable. We have reason to think it is not. More importantly, however, the advocate of the conceptually-indifferent approach may have ready answer to the question of why we should accept a view going against our intuitions: the theory may be empirically useful, allowing us to predict and explain important phenomena. Conceptually-indifferent scientific naturalism is really *pragmatic* naturalism.

Stich (1992) suggests that this is just the kind of answer that an advocate of a fully naturalistic naturalism about the mind should give. He points out that it is parallel to a great many other philosopher's endeavors in relation to existing sciences. Often scientists will use poorly defined or undefined concepts to yield empirical success (he gives examples of "fitness," "grammaticality," and "space-time.") The job of philosophers of science is to examine the scientific use of the concept and make its meaning more explicit, perhaps even to propose improvements. This is exactly the approach a naturalization of the mental might take, by looking to existing notions of "representation" in use by the best cognitive science, describing them and perhaps "patching them up." Stich even points to the work of Cummins (1989) as exemplifying this approach in the philosophy of mind. Cummins admits that he is not trying to analyze any folk psychological concept. Which concept of "representation" we should to explicate, says Cummins, is a question of choosing a theoretical framework and finding the concept of "representation" that plays an explanatory role in it. Cummins says he wishes to explicate the notion of "representation" used in "'orthodox' computational theories of cognition," which "assumes that cognitive systems are automatic interpreted formal systems" (1989, 13). According to Stich (1992, 253), the upshot of this attitude is a pluralism about concepts of "representation":

> [I]f different paradigms within cognitive science use different notions of representation, then there isn't going to be *a* theory of mental representation of the sort we have been discussing. There will be *lots* of theories. Moreover, it makes no sense to ask which of these theories is the right one, since they are not in competition with one another. Each theory aims to characterize a notion of representation exploited in some branch of cognitive science. If different branches of cognitive science use different notions of representation, then there will be a variety of correct accounts of mental representation….I see no reason to suppose that there is a unique correct framework for theories in cognitive science.

Apart from the apparent plurality of concepts of "representation" to be found in science, there is another reason that motivates Stich's pluralism here. When discussing the first two naturalization proposals, we have seen how each is in effect underpinned by a theory of reference. Analytic naturalists generally determine the reference of "belief" by looking exclusively to descriptions associated with the concept, while conceptually-regulated scientific naturalists generally rely on two-factor theories of reference that are compatible with a causal-historical account of reference. But Stich (1996, 37–54) wonders what it would even mean to have a naturalistic theory of reference to begin with. (He asks this question with attention to determining the reference of "belief," not the reference of particular *beliefs,* but we will later see that the questions are interrelated.) He considers that a theory of reference might be an account of our folk semantic intuitions, or a "proto-science" identifying some scientifically useful "word-world relation." The folk semantic account, of course, would involve the same types of problems we have already seen in attempts to naturalize folk psychological intuitions: these intuitions, even when pitched at a generic level, are variable and fallible, and there is evidence suggesting that they are culturally idiosyncratic. The proto-scientific account of reference would be as useful (or not) as the account of representation or belief that we are currently considering. Presumably different scientific research programs could exploit different word-world relations to explain different phenomena. So there would be no single correct theory of "reference": on some accounts, "belief" might *refer,* while on other accounts, the same concept "belief" might fail to *refer** (where "refer*" exploits a concept of "refer" useful for some different purpose). From all of this, Stich concludes that there are no determinate facts to adjudicate between competing theories of reference and determine a single, correct notion of "reference." For this reason, he thinks that looking to reference to determine ontology (what Quine called "semantic ascent" and what he and Bishop call "the flight to reference" (Bishop and Stich 1998)) is a hopeless pursuit. So just as there can be no single correct notion of "reference," there can likewise be no single correct notion of "belief" furnished to us by a single correct notion of "reference."

This pluralism about concepts of "belief" or "representation" is not necessarily a problem for the conceptually indifferent naturalist. If the lesson is that we choose scientific concepts of "representation" for the sake of their explanatory power in a particular domain of research, then provided that we have such a domain of research we are interested in explaining, we should be able to find a relevant notion of "belief" or "representation." Fortunately, we did enter this discussion with a research program: we entered from the domain of epistemology. So the crucial question we must now address is: is there a scientific concept of "belief" or "representation" that will suit the purposes of a naturalized epistemology, one that will yield an understanding of beliefs that can be true or false, beliefs that can be produced by reliable or unreliable processes? And most importantly, will this meet one of the goals mentioned by Kitcher, will it enable us "to understand and improve our most sophisticated performances," i.e., our most advanced scientific beliefs?

Because of the plurality of "representation" concepts that are possible under a conceptually indifferent approach, we obviously cannot examine every one for its conduciveness to epistemological purposes. At best we can examine a few representative cases. A good place to begin is Cummins himself. Stich's imprimatur suggests, at minimum, that his outlook is purely naturalistic. So, does it help us with epistemology?

According to Cummins, representation is a different issue from intentionality. Whereas intentionality concerns the content of thoughts (conceived in terms of belief-desire folk psychology) Cummins is primarily concerned with the sort of representations involved in computational systems (1989, 88). This does not yet mean that computational representation will have no explanatory value in explaining mental content; it is merely a well-understood starting point that may shed light on issues beyond computation. According to this computational view, for a system to function as a representation is simply for its elements to be isomorphic to the objects represented. That is, if there exists an interpretation mapping elements of one system on a one-to-one basis onto elements of another, the first system *represents* the second system. A simple example is an adding machine whose buttons, internal states, and display serve to represent the argument, function, and output of various mathematical operations. Representation, then, is simply a kind of simulation by one system of another. This concept of "representation" is said to explain a important facts about practices that use representations, like the operation of calculators.

Explaining the operations of calculators is one thing, but what about the theorizing of scientists? At first, it might seem like Cummins' notion of representation is an ideal fit for explaining scientific reasoning. He says that computationalism "embraces a rationalist conception of cognition": a representational system counts as *cognitive* when it serves to facilitate behavior that is "cogent, or warranted, or rational relative to its inputs" (1989, 108). A cognitive system is an "inference engine" that relates propositional contents to other propositional contents. The objects of this kind of computation are symbols, which represent conclusions and premises (109).

There is quickly some trouble, though. If cognition occurs only where there is inference, and if inference is governed by the laws of the special sciences, then where there are no special sciences that postulate inference-facilitating "natural kinds," it seems that there can be no cognition. If there are no special laws of clothing, for example, there might be no cognition about clothing (112). This computationalist theory of cognition, then, will work "only for the cognition of autonomously law-governed domains" (113). Cummins thinks this is unsatisfactory, and a more satisfying account of cognition will require a functionalist specification of various modes of cognition: "cognition will [need to be] what humans do when they solve problems, find their way home, etc." (113). We might stop at this point and wonder how exactly this functionalist specification of cognition is supposed to work. It is not clear if it would involve any of the non-naturalist commitments for which we have attacked other functionalists. (The chapter that describes the proposal does not make clear what kind of functionalism it is—only that it is to be distinguished somehow from causal and conceptual role functionalisms (115–9).)

But this problem of specifying cognition functionally already assumes that the computationalist view of cognition is more advanced than we actually have reason to believe, because it assumes that an isomorphic account of representation can furnish inferences among propositional representations. The problem about "clothing" cognition takes for granted that law-assisted inferences would be unproblematic, but really explaining the possibility of representation of and by scientific law is the greatest challenge for an isomorphic view. The closest Cummins comes to discussing law-based representation is an example concerning Galileo's use of geometrical diagrams to calculate the distance traveled by an accelerating body. There is a clear sense in which the structure of the diagram is isomorphic to the magnitudes of motion it is used to represent (94). But no where is any indication given how to generalize from this example.

Clearly much scientific reasoning does not proceed by the use of diagram, but through the use of mathematical and conceptual representations. If mathematical equations are isomorphic to the systems they model, then this must be so in virtue of our conceptual interpretation of the marks on the page. Consider a simple physical equation such as $f = ma$. Is the equal sign in this equation isomorphic to some aspect of a body under force? No one would say it is. The equation derives from our knowledge of a proportionality relationship: given a constant applied force, the amount of acceleration is inversely proportional to the mass, hence the product of an arbitrary mass and an arbitrary acceleration is constant. The equal sign here represents no single aspect of a body under force, but indicates the fact that any given constant force creates a constant proportionality relationship between a range of values of mass and acceleration. If the marks on paper bear any isomorphism to these facts, it is only in virtue of our understanding of the concepts of varying force, mass and acceleration. Now of course once the equation is understood this way, it can be *used* in a way that is dynamically isomorphic to certain systems: as we change the acceleration variable, the mass variable will change in the same proportion as *acceleration* itself changes with respect to *mass* itself.

Of course, in a later work, Cummins (1996) himself admits and even insists that neither concepts, language, nor knowledge structures in general serve any representational role (131–46). So perhaps a physical equation was never meant to count as an example of a representation in the first place. This makes sense, since the idea that language is isomorphic with reality has gone the way of the *Tractatus*. Perhaps the full story of scientific cognition requires a story about how non-representational devices like equations get connected with other more clearly representational devices, or perhaps certain domains of scientific cognition do not require representation at all (recall that Cummins has distinguished representation from intentionality). Yet it does seem odd to count a geometrical diagram as a scientific representation but not a physical equation. Certainly many scientists *think* that equations represent important physical relationships without being isomorphic to them. And while *understanding* physical equations presupposes many other concepts and abilities, and that given this understanding, equations can be used in an isomorphic way, it still seems that the equation itself represents an abstract fact about physics that is not reducible to any isomorphism. All of this seems disingenuous to a naturalized epistemology that wants to explain the most "sophisticated performances" of scientists.

But perhaps this line of objection defers too much to our pre-theoretic concept of "representation," and does not yet take seriously using that concept as a term of art for a particular explanatory purpose. This is not the only gap between the isomorphic conception of representation and our pre-theoretic concept. By Cummins' own admission, the isomorphic conception of representation furnishes representations on the cheap. That is because any given structure will be isomorphic with a multiplicity of other systems. One consequence of this view is that there is a serious gap between representation and intentionality. Whereas isomorphic representations are "radically non-unique," intentional content is supposed to pick out exactly one state of affairs. To use his examples, intentional content is the kind that makes a thought about Central America just about Central America, and not also about Gödel numbers (1989, 137–8). One implication of this is that isomorphic representations will certainly not support any anti-individualistic notion of "wide content": the states of a representational system, on his view, are individualized entirely by the computational states of the system (116–7). He also doubts that any functional specification of cognition of the kind described above could ever help to produce intentional representations out of non-intentional ones (142–3).

The incompatibility of isomorphic representations with intentional content, particularly with "wide" intentional content, is especially threatening to the endeavors of naturalized epistemologists. In my first chapter, I already mentioned how some naturalists like Kitcher make significant use of causal theories of reference (closely related to theories of wide content), in order to explain the continuity of reference of trans-theoretic terms, in order to answer "pessimistic meta-induction" arguments against the reliability of science. But there are, perhaps, even more ways in which traditional projects in naturalized epistemology depend on wide content. One recent, provocative argument by Majors and Sawyer (2005) even suggests that the notion of wide content is crucial to reliabilist theories of justification, by way of serving to answer one of the most daunting objections faced by the theory. According to their argument, only an externalist reliabilist conception of justification keeps justification truth-linked. Yet an infamous objection to reliabilism brings about reformulations of reliabilism that de-link justification and truth. A twin of ours living in an evil demon possible world might engage in the same reasoning practices as ours, yet be radically mistaken. It seems that since he is reasoning responsibly, he is justified. But then reliability is not necessary for justification. Reliabilism can be saved, say Majors and Sawyer, if true justification is reliability in one's "home world," where the home world is the one in which a subject actually develops, and his contents are individuated in a wide fashion. This allows us to explain how our twin, the victim of the demon, fails to be justified because he fails to have reliable beliefs: even though it may seem that his mental life is the same as ours, in fact it is not—because his mental states are individuated by a different environment. This may be a provocative and controversial use of wide content for a reliabilist epistemology, but it is, no doubt, consistent with a long tradition in naturalized epistemology of making use of externalist or causal theories. (Whether causal theories themselves can be naturalized themselves is, of course, a question for the previous section.)

Waskan (2006) proposes a conceptually indifferent naturalist theory of representation which is similar in many ways to Cummins', but which permits room for wide, intentional content. Waskan assumes that the main purpose of cognitive science is to explain how "we humans are able to behave in a such an unhesitating…and effective manner in the face of even highly novel environmental conditions" (90). In barest outline, this is to be explained by positing a capacity of forethought that permits us to represent the way the world is, and manipulate this representation in order to represent the way we would like it to be (90–1). Another way to think about this is: the purpose of the concept of "representation" is to explain and make predictions about our very ability to explain and predict. Like Cummins, he believes that the kind of representation needed to explain this ability is an isomorphism. But since isomorphism is "cheap," he also tries to specify the kinds of isomorphisms that are relevant to explaining behavior: these are isomorphisms between a subject and a system, which connect to the subject's behavior-guiding mechanisms and permit a subject appropriately related to that system to function successfully in it (96).

Now Waskan argues convincingly that even on this isomorphic conception of representation, the content of representations is still wide. But this does not imply that mental states *themselves* are anti-individualistic. Even if mental content is determined in part by external aspects of the environment, mental content is thereby a relational property of mental states, not constitutive of the identity of the states themselves (80–83). So wide mental content does not threaten *weirdness* of the mental; the mind itself does not "reach out and touch" the world. So mental states themselves can still be described in an individualistic manner, relevant to psychological explanation. What's more, simply because wide mental contents have no immediate causal bearing on subjects in the way that the psychological properties of mental states do does not mean that they have no explanatory value in their own right. Waskan argues that causal impotence does not imply explanatory impotence if *knowledge* of a causally impotent relational property can still furnish predictive inferences. Drawing on an example from Cummins (1996), Waskan tells the story of the Autobot, a small car guided by a slotted card that successfully navigates a maze, even though it does not come into contact with the walls of the maze. Even though the isomorphism between the slotted card and the walls of the maze is merely a relational property, and there is no immediate causal relation between the two of them, *knowing* about this isomorphism still helps us understand why the Autobot is successful (104). It seems that our knowledge of wide content could serve the same explanatory purpose: by using isomorphic representations of our own, we are somehow able to act successfully in the world (105).

Now much of this seems fine to me. I'm sure there is a sense in which isomorphisms can help explain certain kinds of successful behavior. The big question is whether the same conception of representation can account for what Kitcher calls "our most sophisticated performances," our scientific reasoning. Cummins says very little in the attempt to explain the relevance of isomorphic representations to scientific reasoning, but Waskan says a great deal more. In fact I think much of what he says is plausible and compelling. But as I will argue below, I think this plausibility comes from presupposing, from time to time, the possibility of representations that are not themselves capable of being understood as isomorphic.

One domain of advanced thinking that Waskan relates to isomorphic representations is the domain of *non-concrete* representations. It seems difficult, he says, to understand our representations of properties such as being a war criminal, ownership, economic inflation, or electricity, in terms of any kind of pictorial isomorphism. Of course non-concrete value-laden concepts like "war criminal" and "ownership" will be difficult on anyone's theory (hence the entire discipline of value theory). But Waskan thinks that the concepts of "economic inflation" and "electricity"—along with many other non-concrete concepts, might be understood by analogy or metaphor to representations we can depict through pictorial isomorphism. Economic inflation, for instance, is presumably understood by direct analogy to actual, physical inflation. While understanding the causes and effects of electricity does not rely on analogy, understanding "the 'thing' itself" does require analogies to the flow of water through pipes, etc. (139).

Now Waskan mentions an objection to the reliance on analogies and metaphors by Prinz, on the grounds that "metaphors leave remainders" (Prinz 2002, 171–2). That is, to think of two things as alike in one respect is also to think of them as different in others. Flowing electricity is not literally flowing water. The remainder is what makes electricity *electricity*, rather than water. Waskan concedes this, and says that the remainder for "electricity" is to be handled by knowledge of the special causes and effects of electricity, which presumably are not shared by water. Where analogy and metaphor are useful is in understanding electrical phenomena themselves, as apart from their causes and effects. I think this position is entirely appropriate (indeed I make use of the cognitive power of analogy in a later chapter). And Prinz wouldn't seem to disagree, since he says that there is nothing wrong with researching the role of analogies and metaphors in cognition. But Prinz's point is that they "neither fully exhaust our understanding of abstract concepts nor explain how they get linked to their referents in the world" (172). This, presumably, is a point Waskan would agree with, but then the question is: even if knowledge of the causes and effects of electricity is sufficient to distinguish our concept of electricity from our concept of water, where does our knowledge of causes and effects of electricity come from, and how does it get represented? Can it be represented through isomorphs of any kind?[24]

The knowledge of cause and effect that leads us to posit the existence of electricity is knowledge about static charges, about the transfer of charges, about the effects of transferring charges on magnets, etc. How might knowledge about charged physical objects be represented with isomorphs? Understanding the different types of charges involves understanding the results of a complex set of experiments (undertaken by Gilbert, Gray and Dufay) involving different degrees of attraction and repulsion of different types of materials. Now a single isomorphic representation in a physical system is presumably very good at representing another *single* physical system. But how does a single isomorphic representation come to represent the diversity of considerations involved in a concept like "charge"? Presumably we shouldn't be afraid to accept help from other representations that help condense some of this diversity for us, like concepts of materials and actions that help us to report the results of these experiments. But the question about isomorphic representation can arise again for some of these concepts. A single representation may be isomorphic with a piece of amber, for example. But pieces of amber come in a variety of different shapes and sizes. When Waskan said he could explain non-concrete domains of cognition using analogies and metaphors, it seemed that we were about to get an answer. But it turns out that many of the remainders of the analogies and metaphors turn out to be non-concrete themselves. This is not necessarily a problem, provided that these non-concrete remainders can themselves be understood in terms of some further combination of analogies or metaphors and other isomorphic representations. But this implies that *at some point* we must have some isomorphic representations that acquire the ability to represent generalized content

[24] In what follows I limit my answer to how I presume this knowledge is acquired for experts, because I assume that the layman's knowledge and beliefs about scientific topics are either incomplete or parasitic in some way on the experts.

(since analogies and metaphors will not create non-concrete representations without the aid of other representations). The question is: *how* is generality of representation created in the first place?

Probably because of this question, Waskan devotes a separate section to the question of genera, or “universals.” To show how genera might be represented using his resources, Waskan gives the example of pre-algebraic, spatial proofs of the Pythagorean theorem (Waskan, 142–6). These involve first constructing literal squares on the sides of any given right triangle, and then, by a series of manipulations (as if using construction paper), showing that quite literally, the area of the square on the hypotenuse is equal to the area of the sum of the area of the other two squares. Now I have no doubt that many such proofs are possible in geometry, and no doubt that each of the manipulations involved in synthetic proofs could be modeled, and predictions about these manipulations made, using isomorphic representations. My concern is whether such a proof really helps establish any truths about *triangles in general*.

Of course there is an ancient explanation for how such proofs *can* establish a universal principle: as long as we recognize that the size or ratio of the sides has no bearing on the outcome of the proof, we recognize that the outcome is true of all right triangles, no matter their size or shape—the same point made by Berkeley against Locke in defense of nominalism. Waskan notes, of course, the likely objection that isomorphic representations would not necessarily account for this recognition that size and shape do not matter. He concedes this, but notes that it does not call into question the fact that the synthetic spatial proof can prove something about every triangle. It can if we have the recognition that size and shape do not matter, wherever that recognition comes from. I will grant this, but note that it simply raises a further question: where then do we get that recognition that size and shape do not matter, if not from representations? Waskan says it is the “combined effect of knowing that each individual manipulation made over the course of the proof would have had the qualitatively identical outcome no matter the initial lengths of the right triangle’s three sides” (145). I will grant that there are some fairly primitive cognitive mechanisms that enable us to understand each step of the proof without resorting to representations. At certain stages, for example, we need to grasp a triangle as staying the same shape through a rotation, or to grasp how two triangles, if lined up the right way, will form a straight line along one of their common edges. Each of these could be accounted for by perceptual-level, non-conceptual abilities. Even still, it seems that in order to know that these same manipulations would yield the same results for any triangle*,* one would need to be able to imagine any *triangle* going through the same operations. But to recognize these as *triangles*, even though triangles can be very different, would still presuppose an abstract, nonisomorphic representation of triangles.

I think there is a way an advocate of the isomorphism view could respond to this challenge, but let me take a brief digression by mentioning that Waskan does have another recourse available. He can use the appeal to metaphors and analogies to explain the very idea of *generality*. In one section, he makes the provocative suggestion that the very idea of category membership involve analogies to actual physical "containers," or "joints" along which nature is to be "divided" (150). Like all other analogies, however, this one too will have what Prinz calls "remainders." There is obviously an important difference between a particular object's relationship to a potentially infinite number of other particular objects (in the case of category membership) and a particular object's relationship to particular a container. Given these differences, fruitful use of the analogy will require answers to this question: In category membership, in what way does a potentially infinite number of objects have a "boundary" that separates these objects from other sets of objects, and such that some objects are "contained" within this boundary, and others are not? Obviously the boundary and the containment are not literal.

It may be tempting to answer this question by saying that it is not the job of a theory of representation to explain how this analogy is possible. It is enough that we *do* analogize things this way; other cognitive mechanisms besides representation may explain how we do. It may also be tempting to appeal to research in psychology that suggests a psychological mechanism accounting for our ability to regard things as members of categories that from a very early age, children have a tendency to be “psychological essentialists,” who categorize superficially different objects as if they shared “hidden mechanisms” in common. But is “as if” really enough to explain the ability to regard things as members of a category, or simply another way of restating the ability that is to be explained? Indeed, saying that we regard superficially different things as members of the same category simply because we think they share some common hidden mechanism is simply to push the problem to a deeper level, since the question must then be answered what makes it possible for us to regard the mechanisms as the same, particularly when they themselves are likely to be different in subtle and important ways.

Finally, there may be something fundamentally misleading about saying that we can understand category membership by an *analogy* to containers, because to say that we analogize two things is to say that we regard one as similar to the other in some explicitly considered respect. Yet it is precisely the ability to regard two things as similar in a respect that we are in effect trying to explain when we attempt to explain the possibility of thoughts about generality, so we cannot appeal to it in order to explain generality. Only if there is some mechanism that might account for an awareness of similarity that does not involve explicit appeal to the *respect* of the similarity might there be a way out.

I believe that the advocate of isomorphic representations may have a way of explaining the implicit grasp of similarity, at least perhaps for similarities with respect to simple attributes, like shape, for the sake of developing representations like “triangle.” One answer is suggested by Prinz (2002), and I actually make use of this suggestion in a later chapter. Prinz describes how our most basic-level concepts might be grounded in experience through the use of representations he calls “proxytypes.” A proxytype is the product of a “long-term memory network,” a system that dynamically represents a *range* of possibilities through the transformation of a single image into another. We can, he says, “zoom in” on our representation of an object in a continuous manner, or even continuously transform an image of one object into an image of another. This kind of representation allows us to see different objects as similar to each other, as the representation of one is easily reachable by transformation of the representation of another (141–44). This is a model which could, presumably, be applied to the concept "triangle." The differences between right triangle, scalene and isosceles are such that we can easily imagine transforming one into the other, into the next, just by growing and shrinking lines and angles. So perhaps there is a source of generality for the isomorphic representationalist after all, provided that isomorphs are understood *dynamically*. Presuming that proxytypes could form the basis for our first source of generality, we could eventually build on this source and acquire further concepts through analogies and metaphors, synthetic proofs, and other resources.

However, there are only so many things that can be represented using proxytypes, as Prinz is the first to admit (166). Strictly speaking, it is a theory for "basic level concepts," concepts of *objects* like chairs and tables, men and dogs, the kind of concepts children first learn and which research shows involve the maximum level of generality without sacrificing a high level of informational richness (Rosch 1978). (We can think of "triangle" as a basic level concept for shape concepts, though it is doubtful that shape concepts could be learned prior to object concepts.) It is very unlikely that proxytypes could be transformed to account for even one level greater of abstraction. For example, while it is comparatively easy to imagine differently shaped chairs transforming into each other, it is harder to imagine chairs turning into tables or beds or shelves, in order to account for the level of generality comprised by "furniture." This would not be a problem if only the concept "furniture" could be defined in terms of the more basic level concepts, but the only available basic level concepts would be "chair," "table," "bed," "shelf," etc., and a disjunctive definition of "furniture" as "chairs or tables or beds or shelves" would beg all of the relevant questions. The very problem to be solved is how these items of furniture come to be associated with each other, given that it is not as easy to imaginatively transform each into the other.

At this point I want to begin to wrap up, by suggesting that the problem faced by these isomorphic theories of representation is a problem faced by naturalistic theories of representation, in general. Consider that a naturalist might, at this point, abandon epistemology for metaphysics, and say that representations succeed in picking out more general properties in the world just in case they bear some appropriate causal relationship to them, perhaps via a reliable covariance between the property represented and the tokening of the representation, *a la* Fodor. Prinz, for example, seems to consider this solution for the higher-level concepts he can't account for with proxytypes (173, 241–51). Prinz considers a problem associated with causal covariance theories which, I think, represents a number of problems with *naturalistic* theories of intentionality, in general, including the isomorphism theory. This is what Devitt and Sterelny (1999, 79–81) have called the "*qua* problem." Suppose that I am in causal contact with a particular wildebeest (or have a representation that is isomorphic with it). But a wildebeest is also a mammal, an animal, a subspecies of gnu, and the prey of lions. Given that our representation is in contact with (or isomorphic with) this particular wildebeest, what then do we say this representation is a representation of? Which of these categories?

Now Prinz thinks he can solve the *qua* problem using the right kind of nomological covariance theory of content. We can say that our concept refers to *wildebeests* rather than to animals because whereas wildebeests reliably cause tokens of *wildebeest*, animals do only under special occasions. The usual problem with causal covariance theories of this variety is the disjunction problem mentioned in the previous section: what is to prevent the content of "wildebeest" from being "wildebeest *or* bush pig," given that the two can sometimes be mistaken for each other? Prinz dismisses Fodor's asymmetric dependence answer to this problem. Instead he invokes a Drestke-style answer that presupposes that there is a well-defined learning-period for a concept. He avoids the modal problems that usually attend these kind of idealization stories by saying that the content is fixed not by what *would* cause a concept during a learning period, but by "the actual class of things to which the object(s) that caused the original creation of the concept belong" (250). The problem is that this solution to the disjunction problem itself raises its own version of the *qua* problem. Which class is picked out by that "incipient cause"? If it is the class of things that look like wildebeests, then it does nothing to solve the disjunction problem, and only deepens the *qua* problem.

The usual solution to the *qua* problem offered by naturalists is to offer two-factor theories of reference. We have already seen this at work in the semantics of David Chalmers, who appeals to the intuitive dispositions of the "primary intension" to fix a sortal that is affixed to a causal source of reference. But other even more palpably naturalistic theories rely on the same strategy. Stanford and Kitcher (2000), for example, show quite convincingly how the *qua* problem can be overcome provided the proper kinds of background beliefs or knowledge. The present point is that this solution only works if we can take the reference of that background theory—*its* representational content—for granted. And yet it is precisely that kind of content we are attempting to account for: if the *qua* problem applies even to a concept as close to perception as "furniture," the generality of great portions of our background theory will need to be accounted for.

There are, in fact, interesting parallels to be drawn between the problem of cheap isomorphism and the *qua* problem. The first of these problems was, in effect, that an isomorphic representation represented too much: very many things are isomorphic to a single representation. The second is that a particular isomorphic representation does not represent enough: it cannot, by itself, represent general categories, because the particular members of general categories differ too much. Cummins thought that the first problem could not be solved, not even with the addition of further cognitive resources: if something represents in virtue of its isomorphism, nothing about how we use it or the further cognitive resources we bring to bear will *stop* it from being isomorphic to too many things. A parallel point exists for the *qua* problem: if isomorphism between a single representation and a single object fails to account for *any* amount of generality of representation, it is unclear how the addition of further cognitive resources would increase the amount of generality without themselves relying on additional general representations of their own. Indeed, as I have argued, cognitive resources like analogy only seem to do the trick when they presuppose other conceptual representations.

It is interesting, incidentally, that Prinz uses the example of "wildebeest" to illustrate his causal covariance theory. It is plausible that there could be a causal covariance between wildebeests and "wildebeest" representations, but it is plausible because something like the proxytype theory can account for *how* there could be such a covariance: the nearly effortless psychological ability to see wildebeests as similar would lead to this tokening. But when we are talking about higher-level concepts, the explanation for the causal covariance is not as obvious. Different pieces of furniture are very different in shape and size. The psychological channel that might have enabled the covariance, via proxytypes, is not available in this case because of the difficulty of imaginative transformation. So the usual solution is to find some property intrinsic to pieces of furniture—like a function—which is common to all, amidst their many differences. Interestingly, a similar approach is at work in the supervenience views we considered in the earlier section. Supervenience is just a kind of *vertical* covariation of properties, rather than the horizontal kind. Yet properties, treated as causal agents—in either vertical or horizontal covariation theories—are what naturalists like Quine would call metaphysical "creatures of darkness." This is no surprise, since Quine (1953c) thought that reified attributes were just as inten*s*ional as meanings and modalities, having the same difficult identity conditions (e.g., the attribute of exceeding 9 = the attribute of exceeding 9, but the attribute of exceeding the number of planets ≠ the attribute of exceeding 9).

If the *qua* problem is as serious as I suggest, it may not just be a challenge, but an insuperable barrier. For there are those who think that higher-level intentionality can never be naturalized. The *qua* problem points to the fact that at some level of abstraction, understanding representation may require appeal to an irreducible element of intentionality. Indeed the problem may be present even at the beginning of abstraction. I have suggested the proxytypes could account for the most basic recognitions of similarities, but of course we *do not* imagine every possible triangle, even if we *can* imagine every possible triangle. Triangle proxytypes seem to offer at best the potential to represent triangles, but only if we add some kind of wordless order to regard all possible transformations within a range as triangles.

One might respond to the *qua* problem as Cummins suggests some might deal with the problem of "cheap" isomorphism, by indicating that content needs to be individuated functionally. So, for instance, one might say that it is not just the isomorphism to a wildebeest that makes for a representation of a wildebeest *qua* wildebeest, but the particular way in which wildebeests interact with us, how our representation allows us to deal with them, etc. Suppose, for example, that we always defend ourselves against wildebeests in a certain way, but not against other mammals. But this solution is more of an abandonment of theory of representation than it is a improved theory of representation. If it is possible to account for general types of responses to objects just in virtue of properties of the objects, we do not need isomorphic representational middlemen to account for our behavior. This would be a Wittgensteinian use theory of meaning, rather than a representational theory. Naturalistic or not, it is difficult to see how this kind of theory would deliver an account of cognitive content of the sort needed by the naturalized epistemologist.

Conceptually indifferent naturalists may well offer accounts of representation that offer genuine explanatory value without resorting to non-naturalistic assumptions. They may also offer theories of representation that plausibly show how the reference of scientific beliefs might be fixed. The problem is that the useful naturalistic concepts of representation do not seem to account for either the uniqueness or generality of scientific reference. Yet this is what is needed from a conceptually indifferent naturalism. We do not require of it that it satisfy our folk intuitions about representation, but we do require that it serve a useful purpose. Since we came to this discussion looking for a notion of representation that would serve our purposes as naturalized epistemologists, concerned with explaining the "most advanced performances" of scientists, it appears that conceptually indifferent naturalists face quite a challenge.

Conclusion

In the above, I have argued that naturalization proposals for the concept of "belief" (or "representation" or "intentionality") fail to deliver the goods needed by a fully naturalized epistemology. Analytic naturalism fails because of its reliance on conceptual analysis and on the substantive notions of meaning that go along with it. Conceptually regulated naturalism fails not only because of its even more implausible reliance on analysis, but because of its reliance on numerous substantive intensional concepts required to make sense of supervenience, none of which pass traditional naturalist muster. Finally, conceptually indifferent naturalism fails, not because it contradicts naturalist methodology in the way that the first two proposals do, but because it fails to deliver the kind of naturalized belief that naturalized epistemologists, studying the origin and justification of advanced scientific beliefs, need.

At the end of my first section, I noted that the naturalized epistemologists who maintained epistemology's need for a naturalized notion of belief were rarely in the business of doing that naturalization for themselves. Instead they decided to let the philosophers of mind do the job for them. But now we can see that this is one particular division of labor that proved inefficient. Even when naturalization proposals seemed more successful on their own terms (as in the case of conceptually indifferent naturalisms) they do not deliver goods in the proper form needed by the epistemologists. Sometimes division of labor is perilous. Sometimes, if you want to do the job right, you'd better do it yourself. The failure of the optimistic naturalists to do the job themselves, in the end, opens the door for pessimists, like Quine, who certainly *did* do the job for themselves—but with a much different outcome than the optimists had hoped for.

CHAPTER 4

DEFLATIONARY NATURALISM ABOUT BELIEF: THE CASE OF SIMULATION THEORY

Naturalists in the philosophy of mind have typically focused their efforts on questions of the *ontology* of belief, by attempting to show how mental states might either reduce to or supervene on the physical. In the previous chapter, I have shown how a wide range of proposals for naturalizing belief in this manner are not successful.

But Huw Price (2004) has noted that there are two ways to approach naturalistic quandaries about concepts like "belief": as an *object* naturalist, or as a *subject* naturalist. The first takes the primary philosophic question to be whether a given controversial type of entity X exists in the natural world, typically answered by deciding if the term "X" refers; the second takes the primary question to be about ourselves as natural creatures, thus primarily about the status of our use of the *term* "X."

According to Price, the object naturalist approach has independent problems apart from the ones I noted in the previous chapter. In particular, if we find it is impossible to naturalize "belief," and decide it does not refer, we are threatened by the incoherence, discussed by (Boghossian 1990), of saying that "content" has no content. And if we take seriously the arguments of Stich (1996), then insofar as there is no way to naturalize theories of reference, there is no determinate way to say *either* that "belief" refers *or* that it does not.

But Price tells us that the second approach to naturalization, subject naturalism, does not face the problems of the first. It avoids each of the problems faced by object naturalism, insofar as it avoids questions of reference entirely. The main purpose of subject naturalistic philosophy, instead, is to "account for the use of various terms . . . in the lives of natural creatures in a natural environment" (81). So a subject naturalism about "belief" is interested in pragmatic usage of the term "belief," not its reference. So while the first approach to naturalism is an inflationary approach to mentalistic ontology, the second is deflationary.

How, then, would subject naturalism propose to account for our use of the term "belief"? Appropriately enough, one of the first overtly naturalistic philosophers to propose such a project was none other than Quine himself. Quine (1960, 219) sketches a model of belief-attribution that he would later call "empathy":

> [I]n indirect quotation we project ourselves into what, from his remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned…. Casting our real selves thus in unreal roles, we do not generally know how much reality to hold constant….We project ourselves even into what from his behavior we imagine a mouse's state of mind to have been, and dramatize it as a belief, wish, or striving, verbalized as seems relevant and natural to us in the state thus feigned.

In his later work Quine (1992) came to rely on the concept of empathy not only to account for "irreducibly mental" ways of grouping "neural realities" (an endorsement of Davidson's anomalous monism) (71–2), but even as an account of language learning (1992, 42–3; 1995, 89–90). Drawing on Quine's insight, and the work of Davidson (1968), Stephen Stich (1983) endorsed a similar view in his earlier work.

However, critics of Quine, such as Ebbs (1994), have worried that "empathy" is a notoriously subjective-sounding concept for a naturalist like Quine to invoke. Perhaps in light of Price's concept of subject naturalism, however, there is less cause for concern. If we can use natural science to account for the mechanism of empathetic practice, we may be able to naturalize empathy and, thereby, the terminology of the mental. This possibility has occurred to several philosophers writing prior to Price's discussion of subject naturalism. Picardi (2000, 131–2), for example, notes that literature on the debate between "simulation" and "theory-theory" approaches to folk psychology might help naturalize Quinean empathy . Picardi thinks that neither side of this debate could satisfy Quine's "behaviorist strictures." But I will argue that one of these options—"simulation" theory—has more affinity to Quine's behaviorism than Picardi suspects. Specifically, I will show that the "radical" simulation theory of Robert M. Gordon (1995a) offers the best promise for a thoroughgoing *subject* naturalization of "belief."

I will begin this chapter by presenting the nature of the dispute between advocates of theory-theory and advocates of simulation theory. I will show why simulation theory, rather than theory-theory, offers the best chance at a subject naturalization of "belief," and why Gordon's version of simulation theory, in particular, is the best-suited version to accomplish this. I will then examine two leading objections to Gordon's simulation theory—the problems of pretense and adjustment—and attempt to defend Gordon against them. In the end, however, I will argue that the best case for Gordon's view runs into a new problem. After presenting this problem, I will discuss some important philosophic lessons learned from an encounter with simulation theory, including the consequences of its failure to serve as a naturalization proposal vis-à-vis naturalized epistemology.

Theory-theory vs. simulation theory

The topic of folk psychology—the question of how human beings understand the minds and thoughts of others—is of considerable relevance in several areas of philosophy. In epistemology, the question of whether we must understand others beliefs as generally coherent and rational in order to interpret individual beliefs is relevant to arguments against skepticism (such as those advanced by Davidson). In philosophy of mind, the dispute between realism and irrealism about the mental often turns on the question of the meaning of mentalistic concepts. If belief-desire psychology, for example, is held to be largely false, it is thought by some that the theoretical concepts of "belief" and "desire" might not refer, with the consequence of irrealism.

The dominant theory of folk psychology for the last century has come to be called the "theory-theory." According to the theory-theory, one person's attribution of a mental state to another essentially involves conceptual representations of that state, usually via the grasp of inferential relations between a theoretical entity and its *explananda*. The theory-theory has been widespread: it is the view of folk psychology presupposed by the classical position that we infer the existence of mental states by analogy to our own introspected states, but also by the more recent socio-linguistic

(Sellarsian) view that claims the reverse: that understanding our own mental states is itself a byproduct of a theory positing states that explain the behavior of others.

In recent years, an alternative to theory-theory has emerged, one which could have important implications for philosophical positions that turn on questions of folk psychology. “Simulation theory” is the name for a family of views formulated independently by Robert Gordon (1995a) and Jane Heal (1995), and subsequently developed by Alvin Goldman (1995; 2006) and others. Simulation theory claims that the more primitive attributions of mental states involve no new representations, but only the use of first-order (non-mentalistic) descriptive abilities, imaginative projection, and decision-making capabilites.[25]

To consider a simplified description of the process simulation theory envisions, suppose, for example, that we want to describe Maxi’s state of mind after he opens the drawer and discovers that the chocolate is missing. Rather than drawing on a theory relating perceptual inputs to behavioral outputs, we simply imagine ourselves in Maxi’s position. We see ourselves opening the drawer, then we feel disappointed, and then perhaps we decide to say to ourselves that the drawer is empty. Of course we make this decision “offline” and do not act on it. Drawing on our simulation, we can now ascribe the associated feelings and beliefs to Maxi, and perhaps predict his behavior.

In this chapter, I wish to focus on the “radical simulation” theory of Robert Gordon, because I take it to be the most consistent departure from theory-theory, and therefore a good test case for evaluating the difference between the simulation and theory paradigms. Other versions of simulation theory, such as Alvin Goldman’s, for example, do not attempt to account for the *meaning* of mentalistic concepts from the ground up in the way that Gordon’s theory does. Goldman presupposes,

[25] In fact a useful way of understanding the difference between theory-theory and simulation theory is just to hold that the former thinks that decision-making capacities are separate from folk psychological prediction mechanisms, whereas the latter thinks prediction capacities are just a part of our decision-making capacities (Davies and Stone 2001). See also Nichols and Stich (2003). At least according to Gordon (1995c) and Goldman (1995), this means that the question of simulation vs. theory can be settled empirically: one need only show that the same neurological mechanisms involved in decision-making (such as those involved in motor skills) are components of systems involved in the response to others (such as the “mirror neurons” activated in response to the perception of other’s action) (Decety 2002).

for example, that the ability to engage in simulation requires a prior grasp of certain basic mentalistic concepts, which are needed to describe the instrospective results of one's own simulation and ascribe them to others, and also to assign the "inputs" needed to adjust a simulation to approximate the differing conditions of the target of simulation. Whether or not Goldman's presuppositions are acceptable on their own terms, they certainly make it difficult to use his version of simulation theory as a subject-naturalization of belief. Even though Goldman's own account of naturalized epistemology is in dire need of a naturalization of belief (see chapter 1), he is not interested in using simulation theory to naturalize mentalistic concepts. He simply wants to explain our application of these concepts to other people without the need to invoke overcomplicated theory-theoretic generalizations:

> In the very posing of my question—how does the interpreter arrive at attributions of propositional attitudes (and other mental states)—I assume that the attributor herself has contentful states, at least beliefs. Since I am not trying to prove (to the skeptic) that there is content, this is not circular or question-begging. I assume as background that the interpreter has beliefs, and I inquire into a distinctive subset of them, viz., beliefs concerning mental states. (2000, 75)
>
> As far as the interpretation strategy is concerned, it indeed appears that if the simulation theory is correct, the interpretation strategy is fruitless. One cannot extract criteria of mentalistic ascription from the practice of interpersonal interpretation if that practice rests on a prior and independent understanding of mentalistic notions. (2000, 94).

Gordon's approach, on the other hand, is "radical" because it (1995c; 1996) denies that either introspection or a prior grasp of mental concepts is needed to simulate or ascribe mental states. According to Gordon, simulation only requires the imaginative projection of oneself into the *situation* of the other. To state the contents of one's own beliefs, one need only ask and answer first-order questions, and append the answers to expressions of belief. One need not ask oneself "Where do I believe the chocolate is?" but simply "Where is the chocolate?" and give the answer, "The chocolate is in the drawer." One then engages in an "ascent routine" in which one appends an "I believe" construction to this result. Therefore all that is needed to attribute beliefs to *others* is to engage in this ascent routine while *in the context of simulating the other* (appending “he believes” instead of “I believe” to the result).

Now it may well be that a hybrid of simulation and theory is needed to understand our mental concepts in all their richness, in which case an approach like Goldman’s may be appropriate. But to get to the point of accepting that, it is necessary to see where a purely simulation-based approach breaks down, and to see that, it is best to examine Gordon’s theory. Furthermore, since Gordon’s “ascent routine” account of simulation does not require any appeal to introspection, it is also the version of simulation theory most likely to find favor with naturalists in the philosophy of mind.

In this chapter, therefore, I will examine Gordon’s theory exclusively and examine some leading objections to it. I will begin by examining some of the experimental evidence that motivates simulation theory, but which also poses preliminary challenges to it. After showing how simulation theory approaches that evidence, I will proceed to examine a serious problem that arises for Gordon’s theory, the problem of adjustment, and speculate about how he could address it. In the final section, however, I will argue that Gordon’s best solution to the problem of adjustment faces a new problem. After presenting this problem and indicating why I don’t think Gordon can solve it, I will present an alternative non-simulation explanation, one which is, nonetheless, not the same as the traditional theory-theoretic explanation.

Preliminary challenges from false belief task evidence

In the widely-replicated “false belief task” experiment, researchers have children watch Maxi (a puppet) place his chocolate in a drawer and leave the room (Wimmer & Perner 1983). After he leaves, his mother moves the chocolate to another location, and then Maxi returns. The children are asked where Maxi will look for the chocolate. Five-year olds typically say Maxi will look where he last saw the chocolate: in the drawer. But three to four-year olds answer that Maxi will look where *they* know the chocolate to have been moved by the mother, indicating some difficulty in understanding or ability to express Maxi's false belief.

This pattern in the false belief task is very often taken as *prima facie* evidence in support of the theory-theory. Early failure at the task seems to suggest a *conceptual* deficit which is remedied by the child's development of a new "theory." In recent years, the relevance of the false belief task evidence has fallen into some doubt, particularly as experimentalists have attempted to simplify the task in a way that controls for difficulties children may have with its purely verbal aspects (Bloom and German 2000; Onishi and Baillargeon 2005). Nevertheless, the view that the task suggests some important conceptual development remains paradigmatic, as recent meta-analyses controlling for task-performance difficulties continue to show a pattern of development (Wellman et al. 2001), and as doubts linger about the significance of the early competence experiments (Perner and Ruffman 2005).

In any case, although recent simulation theorists such as Goldman (2006) draw on recent experimental developments to challenge the significance of the false belief task, at least at an early stage, Gordon (1995a) thought that the false belief task results could be equally or even better explained by simulation theory. He suggested that theory-theory could not explain the development through the acquisition of a new theory, because prior to that acquisition, a child would simply be *unable* to make predictions about human behavior, rather than what actually happens: making *unreliable* predictions (69–70). Now this objection is probably addressed by the theory-theorist's contention that children may have an early theory accounting for the possibility of unreliable predictions: instead of thinking of the mind as a representational device (as they do later), young children could think of the mind as a "copying" device, on which only real objects impress themselves.

But even if theory-theory could explain early false predictions adequately, Gordon (1995a, 70) could still respond that simulation theory offered a *better* explanation:

> Suppose…that the child of four develops the ability to make assertions, to state something as fact, *within the context of practical simulation*. That would give her the

> capacity to overcome an initial egocentric limitation to the actual facts (i.e., as *she* sees them). One would expect a change of just the sort we find in these experiments.

In other words, prior to passing the "false belief task," the child only has the ability to engage in the ascent routine without being able to project himself into the position of the other. From this perspective, the simulation theory explanation of egocentric error seems *simpler* than theory-theory's. Theory-theory has to posit a special children's theory of mind just to account for these errors, whereas simulation theory can simply draw on one's existing descriptive, and decision-making abilities, whose existence no one would dispute. Of course what accounts for the alleged deficit in the imaginative capacity is a matter of some controversy, but it is also not controversial that we eventually develop this capacity, and plausible that young children don't start out with everything.

The important question, then, is whether simulation theory can offer an adequate account of which particular imaginative capacity develops in such a way as to permit the eventual passing of the false belief task. And hopefully, the simulation theorist can offer this account in such a way that retains its simplicity advantage over the theory-theory. Positing unlikely and complicated imaginative capacities could counteract whatever simplicity advantage simulation theory has because of its reliance on existing descriptive and decision-making capacities.

Gordon's musings on this subject are only barely suggestive. He says that if, in the context of simulation, a child is led by evidence from the situation of the simulation to make assertions that conflict with assertions from one's "home" perspective, she will be making *de facto* "motivated attributions of false belief" (1995a, 61). There are questions we should ask immediately about this proposal. First, what kind of evidence is supposed to motivated assertions incompatible with one's home beliefs (hereafter, "home-incompatible assertions? The answer to the first question is simple enough. Presumably a child is motivated to make a new and different assertion because of some anomalous action on the part of the simulation target. The child may see Maxi headed to the "wrong" place: the place where the child knows there is no chocolate. Frustrated by the anomaly, the child looks for an explanation. Of course this would not yet explain children's ability to *predict* that Maxi will go to the wrong place (since in the case of a prediction, the anomalous action has not yet occurred). But perhaps the ability to make such predictions is strictly theoretical, deriving from a simulation-based capacity to give these situation-based explanations.

Next, however, we should ask what imaginative capacities are available to enable this sudden development in the ability to make home-incompatible assertions—and in a way that would allow for the development of the genuine understanding of false beliefs. This is an important question because if the situational evidence that motivates the child to make a home-incompatible assertion is merely the target's anomalous behavior, there is at least one explanation a child could invoke that would explain the anomaly by way of an incompatible assertion, but which would not be the same as a false belief explanation. Supposing that Maxi is seen heading to the "wrong" place (where the chocolate *used* to be), the child could just as easily explain this by making an assertion about what Maxi is *pretending*: that the chocolate is in the drawer.

Now some psychologists do see pretense itself as involving some grasp of representational states in others, but it seems there are some important differences between pretense and false belief that would call this into question. Paul Bernier (2002) thinks that not just *any* mental process leading to home-incompatible assertions suffices for the ascription of beliefs that the simulator could genuinely comprehend as being false. To genuinely comprehend a mental state as a belief, Bernier notes, the simulator needs to know something about the function or aim of the state. Beliefs, in particular, are mental states that "aim at describing one and the same objective reality" (42). [26] A mental state counts as a false belief only insofar as it fails in this aim. But of course pretense does *not* aim at describing objective reality, and so a motivated attribution of a pretense which fails to correspond to reality is not yet the attribution of a representational state: pretense does not misrepresent reality because it is not even *trying* to represent reality.

[26] Jacob (2002, 100-3) seems to be making a similar criticism of Gordon.

In fact, one study by Perner et al. (1994) suggests that whereas most 3-year olds can discriminate between cases of knowledge and cases of pretense, they cannot discriminate between pretense and false belief. Young children surely display that they know their make-believe is not real at an early age (Perner 1991, 66–7, 76–78).[27] Children first acquire the ability to engage in knowing pretense as early as their second year, fully two years before passing the false belief task (Perner 2000, 385). So even if there is some rationale according to which understanding pretense involves understanding a representational state, it is not the *kind* of representational state the grasping of which is necessary to explain the ability to pass the false belief task. If we are trying to explain that ability, then, by some mechanism that leads to home-incompatible assertions, we had better specify that mechanism enough so that it explains more than pretense-attribution can explain.

Gordon, it seems, is aware of this problem, and recognizes that the mere ability to make home-incompatible assertions is not enough to count as the genuine comprehension of false-beliefs. He tells us that the route to a more "sophisticated" understanding of "belief" comes through a novel kind of simulation within a simulation:

> To see her own present beliefs as distinguishable from the facts she will have to simulate another for whom the facts are different—or, more broadly, adopt a perspective from which the facts are different, whether this perspective is occupied by a real person or not—and then, from the alien perspective, *simulate herself.* (1995c, 61–2)

This, however, does not yet seem to give us what we want, because the main question is how, in the first place, a child *can* come to "simulate another for whom the facts are different." The child cannot do it simply by engaging in the "ascent routine" on the basis of the child's own apprehension of the facts, and it also cannot be done through ascription of pretense to the other.

Now the second, broader suggestion, to "adopt a perspective from which the facts are different, whether this perspective is occupied by a real person or not," is slightly more suggestive, if only because it is broader. If, by "different facts," Gordon means "*incompatible* beliefs about the

[27] Children first acquire the ability to engage in knowing pretense as early as their second year, fully two years before passing the false belief task (Perner 2000, 385).

facts," then we are right back where we started. But perhaps Gordon means a different kind of difference, difference without incompatibility. Perhaps the child can simply imagine that he has more or less information about a given circumstance, and in some way the possibility of genuine false-belief attribution emerges out of this. The question of how a child might be able to do this, using existing imaginative capacities, is the problem we will examine in the next section: what Daniel Weiskopf (2005) has called "the problem of adjustment."

The problem of adjustment

To understand how existing imaginative capacities could be used to "adopt a perspective from which the facts are different" in the second sense I've described above, let's first present an example of Gordon's that is meant to warm us up to a kind of simulation explanation that could successfully generate the genuine comprehension of false beliefs.

Imagine that Sam (the simulator) and Tom (the target of simulation) are hiking a mountain trail together.[28] Suddenly Tom yells "Go back!" and Sam follows him, looking over his shoulder. Sam looks over his shoulder, looking for "menacing, frightening things" up the trail. Sam spots a grizzly bear. By looking at the same environment as Tom, and experiencing the same fear, Sam simulates Tom and understands why Tom runs: Tom runs *because of the bear.* Even though this is not yet anything like false belief explanation, it involves a kind of simulation that is more than attributing a pretense, as regards both its cause and effects: it is caused by the apprehension of an anomaly in Tom's behavior, and it has the effect of enabling further predictions about Tom's behavior (that he will continue to run away to safety once he is out of the bear's range), which predictions Sam can use to his own benefit (by following Tom).

Gordon calls this kind of simulation "total projection," because it involves the use of the total set of Sam's own knowledge, beliefs, and desires to understand Tom. No "adjustment" is needed,

[28] The names here are mine, not Gordon's.

since Sam is, for all practical purposes, in the exact same position as Tom. Granting for the moment, then, that "total projection" has explanatory power, we must consider whether simulation of situations more closely resembling false belief explanation have even more power. Gordon (1995b, 106) considers these situations to be ones in which "total projection" must be "patched". In these situations, the reliability of total projection fails, which failure itself constitutes a new anomaly necessitating a new explanatory tactic that will restore reliability. The *problem* of adjustment is just the problem of how total projection can be patched to provide the full extent of explanatory power usually attributed to false belief explanation.[29]

The failure of total projection is seen by theory-theorists as evidence for why interpreters need theory, not simulation, to understand the behavior of others. But Gordon (1995b) suggests ways in which total projection can be patched without resorting to theory. For example, if Sam sees Tom run from the grizzly bear, he may not at first understand the action *if the bear is heading towards Tom but away from Sam*. Since Sam is not in any danger himself, he cannot immediately empathize with Tom's action. The needed adjustment is easy enough: Sam imagines himself looking at the bear from Tom's perspective. Now he imagines the bear coming towards *him*, feels the fear, and can explain Tom's action as stemming from the oncoming bear (107). This almost functions like a false belief explanation, because "there is a bear coming towards me" is *false* from Sam's perspective. What he discovers is that it is not false from Tom's. But there is a question we can ask here and return to later: *how* is Sam prompted to consider what things look like from Tom's perspective?

To begin to understand why that is an important question to ask, consider a modified version of another one of Gordon's examples, this time one that involves Tom's being "epistemically handicapped." This time Sam sees the grizzly in Tom's path, twenty feet ahead, and expects him to run (having used the patching described above). But Tom does not run: he keeps walking towards the

[29] The problem of adjustment involves more than just adjusting to different informational situations. The simulator may also have different evaluations than the other being interpreted. See Weiskopf (2005) for more discussion of Gordon's attempted solution to this version of the problem, and the problems involved in his solution.

bear. As it happens Tom is myopic and sees only a blur ten feet in the distance.[30] Gordon would say that Sam can also explain this action by simulating Tom's blurry vision. Now Sam imagines that he doesn't see the bear in front of him, and doesn't want to run. This helps him grasp that Tom doesn't run because he doesn't see the bear. In another example, Sam sees Tom run from something he *shouldn't* run from: a large Newfoundland (the affable dog breed). Gordon proposes that Sam can explain the behavior by imagining the dog to be from some actually dangerous breed, or imagining the dog to look like a grizzly.

Gordon's strategy here is clearly inadequate. The first problem is that there is no obvious mechanism accounting for why Sam considers these particular hypotheses—or, more precisely, the particular pretend scenarios: to let his vision go blurry, or to confuse a Newfoundland with another dog or with a grizzly. There are many hypotheses that could also account for the behavior exhibited by Tom, and Gordon never tells us how any mechanism is supposed to pick one. Presumably if Sam is an adult, he has a great deal of theoretical knowledge that could be invoked, but at first glance it appears to be knowledge about Tom's mental states—and it is precisely this knowledge that we are not relying on in our attempt to explain how a *child* can come to attribute false beliefs in the first place. This is not a problem that attends the simpler types of simulation. Elsewhere Gordon reminds us of evidence documenting "imitative mechanisms" that lead young infants follow the gaze of others (joint attention), and imitate the facial expressions of others (emotional contagion). But he gives no evidence concerning mechanisms that could somehow lead us to lock on and simulate malfunctions or confusions in a putatively *internal* state of the other.

Perhaps this problem is an artifact of poorly chosen examples on Gordon's part. Consider a simpler example involving the hikers. Suppose Sam is far enough away to see a grizzly that Tom

[30] Gordon's original example involves the myopic hiker turning back anyway for an inexplicable reason. His point is that if the hiker is myopic, the ordinary explanation will not work, as he will not see the bear. The trouble with the example is that it is supposed to illustrate how simulating blurry vision could figure in an explanation, but it only succeeds in showing how a putative explanation can be dismissed. I have modified it so that the blurry vision actually explains the anamolous action.

cannot, because a boulder blocks Tom's line of sight. In this case, Sam could respond to the situation by placing himself in Tom's location, and then, by exploiting his joint attention imitative mechanism, look at the boulder rather than the bear. Feeling no obvious fear about the boulder, he would imagine himself walking forward rather than running away. This type of simulation seems to require no recourse to theory of mind, and yet provides something mimicking a false belief explanation: Tom doesn't believe there is a bear in his path, because he believes there is only a boulder in his path. This looks much simpler than Gordon's idea of the simulator as the young hypothesis-tester. Perhaps, by building on simple explanations like this, more complicated simulations are possible.

Unfortunately, the improved clarity of the example also brings a deeper version of the problem of adjustment to the forefront, one that applies even to the simplest examples of simulation that Gordon has offered us. We need to think more about the *motivation* of the simulator. Taking for granted that Sam is surprised by Tom's movement towards the bear, and is prompted to search for an explanation, what motivates him to imagine himself in Tom's *location* in the first place? This is the same question we asked even before we considered cases of "epistemic handicap," as when the bear was simply headed towards Tom rather than Sam. Now, appealing only to the joint attention mechanism, we can imagine why Sam would look at the rock. But from Sam's vantage point, he can see both the rock and the bear, and he isn't inclined to walk towards the bear as Tom is. So there are mechanisms for attending to what Sam is attending to, but attending to the same thing from different perspectives does not yield the same beliefs. There are mechanisms for imagining what it would look like from Tom's perspective, but none that account for the *motivation* for looking from Tom's particular perspective, as opposed to so many others. There are mechanisms for imitating the other, e.g., his facial expression, but none that account for the desire to imitate his location.

One is tempted to say, perhaps, that of course that location is relevant, because information about spatial location is relevant to how Tom perceives the world. But the question is: isn't this a piece of theoretical knowledge about Tom's mental states, rather than a product of simulation? Perhaps advocates of the simulation theory who are willing to tolerate hybridized versions of it can admit as much, but it would seem that Gordon cannot—not if he wants to show how a primitive grasp of false beliefs is possible using nothing but simulation-based resources. In the next section, I will explore what I believe might account for the motivation to consider another's situation, and why recognizing it undermines simulation theory.

The problem of epistemological adjustment and the complexity of simulation theory

It would seem that the simplest explanation for wanting to imagine oneself in another's position is that Sam knows something about the connection between action and *knowledge*, specifically perceptual knowledge. He knows that people move towards good things they can see, and that they run away from the dangerous things they can see. So in order to determine what Tom can see from where he is, he simulates that position to explain Tom's action.

Gordon himself constantly insists that simulation involves the *implicit* attribution of knowledge to the other. In total projection, it involves projecting all of one's own knowledge to the other (1995b, 103). In patched projection, it may involve pinpointing "relevant ignorance" (1995b, 110–1). In saying this, however, his emphasis is on "implicit": knowledge attribution, which he thinks to be implicit in just the same way that belief attribution is supposed to be. Presumably simulators should be able to attribute it without the need for any explicit theory or concept of knowledge.

But Gordon's view neglects the fact that there is good reason to think that children form explicit beliefs about epistemic facts *before* they pass the false belief test. So there is good reason to think that any simulation with the power to render the same explanations as explicit false belief explanation *can* certainly draw on these explicit beliefs.

First we should say something about the formation of the concept "knowledge". It seems children make explicit mention of it at a very early age. Developmental psychologists have observed children using "know" as early as 15 months (Bretherton, et al. 1981). According to one study "know" is the most frequently used mental term between 2 and 3 years of age (48% of all mental verbs in one child), and of these, “I don’t know” is the most common construction among uses of “know” (62% of them, for the same child). Of course at the earliest stage children use the term ‘know’ merely as a conversational marker, in idiomatic phrases like “Know what?” or “You know?” Still, it is estimated that perhaps 12% of the earliest uses are in the context of a genuine assertion of correspondence with facts (Shatz et al. 1983) This is particularly clear when children don’t simply say “I know,” but contrast their usage with something they don’t know, or talk about how they didn’t used to know something, but now they do (Hogrefe et al. 1986). Clearly something like an “ascent routine” can facilitate the transition between idiomatic and genuine referential uses of “know.” It’s particularly interesting that in longitudinal studies, children regularly draw this knowledge-ignorance contrast in natural conversation months before they begin to draw belief-reality contrasts of the kind associated with passing the false belief test (Bartsch and Wellman 1995, 120).

Second, having an explicit concept of “know” enables children to develop explicit knowledge about the sources of knowledge. There is certainly a period of development in which children have *merely* implicit knowledge of the connection between perception and knowledge. In one study, children are asked which of two characters knows what is in a box: the one who has looked, or the one who has not. By the age of three, most children answer the question correctly (Pratt and Bryant, 1990). But in another study, while three year-olds can correctly identify the lookers as the knowers, the majority are not able to explain how it is that lookers know (they cannot explain where knowledge comes from). Only later, between the ages of four and five, do they acquire the ability to give explicit explanations (Wimmer et al. 1988; Gopnik and Graf 1988). Particularly interesting is a very recent study on the specific relationship between children’s ability to offer epistemic explanations and their ability to pass the false belief test. The results suggest that while there is an initial period in which children fail to offer epistemic explanations and fail to pass the false belief test, there is a transitional period in which they can offer epistemic explanations without yet passing the false belief test (Burr and Hofer 2002). So there seem to be three stages in the development of explicit knowledge: 1) drawing the knowledge-ignorance contrast, 2) identifying perceivers as knowers, and 3) drawing the knowledge-false belief contrast.[31]

It would be rash to conclude that simply because explicit epistemological theory precedes explicit "theory of mind," therefore children must be able to pass false belief tests *because* of development in their own epistemological theories. But this conclusion becomes more tempting if we can offer a hypothesis accounting for the conceptual process that would account for it. Recall that the fundamental distinction between understanding false-belief-motivated action and pretense-motivated action is that only the first involves the understanding of incompatible propositions aimed at the same reality. A child will see no contradiction between his own belief that there is no chocolate in the drawer, and Maxi's pretending that there *is* chocolate in a drawer.[32] But if an (older) child thinks that Maxi is in error, he thinks Maxi believes there is chocolate in the drawer, even though the child knows there isn't. This is because he understands Maxi's attitude as *aiming at reality*, and failing in its aim. In short, "believe" involves a *normative* element that "pretense" does not. I want to suggest that it is the child's *explicit* understanding of the causal connection between perception and knowledge that makes possible understanding the normativity of belief.

It is important that children begin with the contrast between knowledge and ignorance. They understand what they have (knowledge) and what they don't, and their naturally curiosity pushes them to get "more" of what they don't have. So knowledge is a goal of theirs: the question is, what is the

[31] There is of course a legitimate question about whether some of this explicit knowledge about the sources of knowledge might itself result from simulation. I have already noted how a child may form that concept through a kind of ascent routine. And there is possibly some connection between a child's understanding of other perceivers as knowers, and joint attention mechanisms, which are doubtless involved in seeing where others are looking. But it is not at all clear why simulation would be necessary for any of this. Children explicitly understand what knowledge is by reference to their own mental states. They surely also connect their own perceiving to their own knowing. I can imagine that simulation may play a role in discovering that other people know. They might notice that others have eyes, too, and consider themselves in the place of others—and conclude that others must know of particular things. But this would not account for the *concept* "know" or any *principled* knowledge about the sources of knowledge; it would only account for particular knowledge about who knows what. What's more, it's not at all clear why a simple argument by analogy, rather than simulation, would not account even for this particular knowledge.

[32] Indeed, we should wonder if pretense even requires any attitude towards a proposition at all.

means? This they begin to realize when they understand that knowledge originates in perception. Along with this comes an understanding of the *limitations* of our epistemic faculties: the understanding that one can only see one side of an object, a limited stretch of the environment, etc. By grasping these limitations, the child comes to see that knowledge is not just a goal, but a goal that must be achieved through effort–and that seekers of knowledge can fail in their quest. It is also plausible that learning about the limitations of perception, in particular, can help a child to understand how appearances can be misleading—a crucial point needed to understand how one can have a mental state that functions like knowledge even when it is not knowledge. Children, indeed, grasp the distinction between perceptual appearance and reality before they pass the false belief test (Gopnik and Astington, 1988, 34.) Once a child grasps that certain of his own mental states have functioned like knowledge in the past (e.g., had the same types of origins and applications) even when they later turned out to work more like ignorance (because they yielded unsuccessful predictions), he needs a concept to denote these states to explain unsuccessful actions resulting from them. This concept not only offers the aforementioned explanatory value, but also does so in a way that accounts for the difference between false belief and pretense. The child can now see false beliefs as aiming at reality in a way that pretense does not.

If this hypothesis is correct, then an interpreter does not need simulation to grasp the possibility of false beliefs. Of course supporters of the simulation theory do not argue that children *need* simulation to understand the concept of belief. They merely claim that simulation theory offers a simpler explanation of where that understanding comes from than theory-theory does. First of all, as I have shown, they claim that simulation explanations can draw on existing descriptive, decision-making and imaginative abilities, without the need to posit special childhood theories. Second, the *content* of the theories children would need to explain behavior be hopelessly complex in comparison to simulation theory. Gordon argues that if the folk have to rely on behavioral generalizations to explain and predict behavior, they will at best be able to generalize about "typical" behavior patterns, but at the same time they would have to allow for countless *ceteris paribus* clauses. Gordon (1995a) certainly allows that behavioral generalizations are sometimes needed, but insists that the simplest way to understand how *ceteris paribus* clauses are filled in is by using them in the context of simulation, such that one can use one's own practical reasoning skills to predict the best course of action in exceptional circumstances (67).

But if we allow that interpreters also have epistemological generalizations about their targets—i.e., explicit knowledge about the connection between perception, knowledge and action—simulation is no longer needed to simplify the application of these existing behavioral generalizations. Consider how the use of simulation might be used to simplify the application of a behavioral generalization:

S1. People run from danger (*behavioral generalization*).
S2. Tom is approaching a dangerous bear (*anomalous observation*).
S3. I only act to avoid danger that is present (*behavioral generalization*)
S4. There is no dangerous bear present [said from Tom's perspective] (*simulation*)
S5. Tom is walking towards the bear because he believes: there is no dangerous bear(*new explanation resulting from the ascent routine*).[33]

But now consider how the attributor's practical knowledge can be used just as easily to inform the application of a behavioral generalization to the situation, without the need to have countless *ceteris paribus* clauses:

T1 People run from danger (*behavioral generalization*).
T2. Tom is approaching a dangerous bear (*anomalous observation*).
T3. People only act to avoid dangers they know about (*epistemological generalization*).
T4. Tom does not know about the bear which is not in his line of sight (*epistemic observation*).
T5. Tom is walking towards the bear because he believes he is safe (*new explanation*).[34]

[33] Now it is possible that the epistemological generalization in (3) is a bit more specific than we would expect a young child to possess. But it is a conclusion reached easily from "people act on what they know," combined with "people act to avoid dangers." The observation in (4) is of course an application of another epistemological generalization, "people cannot see around solid objects." There is no reason to think a younger Sam would need to consider all of these more general generalizations explicitly at the moment in order to understand Tom, but it is plausible that he would have them stocked in his background knowledge explicitly at some point in the past.
[34] Now it is possible that the epistemological generalization in (3) is a bit more specific than we would expect a young child to possess. But it is a conclusion reached easily from "people act on what they know," combined with "people act to avoid dangers." The observation in (4) is of course an application of another epistemological generalization, "people cannot see around solid objects." There is no reason to think a younger Sam would need

Each of these processes involves parallel steps, and works by applying one's own practical knowledge to a situation (thereby obviating the memorization of complex exceptions). The difference is that while all of the steps in the second process are likely to be known or knowable by young children, step S4 in the first is knowable only if one adds the complexity of a special mechanism that motivates the simulator to envision himself in Tom's position, a mechanism that is not obviously accounted for by any currently known cognitive mechanisms (even though there are known mechanisms involving imagination, joint attention, and facial imitation).

So if simulation is not needed to simplify the application of behavioral generalizations—and if simulation mechanisms require positing new mental mechanisms to explain the motivation to imagine oneself in another's position—simulation theory loses its chief explanatory advantage: it's simplicity. It is not simpler than the explanation involving explicit epistemological generalizations, and may in fact be more complex. The use of explicit epistemological generalizations requires no new cognitive mechanism, just the usual kind: background conceptual generalizations of a type which children are likely to have.

It should be noted that while the style of explanation I have present above is not a version of radical simulation theory, this does not mean that I think simulation never has a role to play in folk psychology. My point about how epistemological generalizations can be used to offer theoretical explanations can apply equally well to simulation: some simulation may be very important in explaining action, but it still requires these epistemological generalizations, and these involve concepts of mental states. The upshot is that any account that attempts to derive the possibility of genuine false-belief attribution and comprehension from purely non-conceptual simulation—like Gordon's "radical" simulation—is doomed. No matter what we say about the relevance of simulation, it must be supplemented by other types of "mindreading." Some kind of hybridized account is necessary. My

---

to consider all of these more general generalizations explicitly at the moment in order to understand Tom, but it is plausible that he would have them stocked in his background knowledge explicitly at some point in the past.

suspicious is that this hybrid account does not require a *prominent* role for simulation, but that is another matter.

What's more, I think that rejecting the possibility that simulation is *basic* to the grasp of mental concepts does not mean that the content of all mental concepts is itself "theoretical," at least in the functionalist sense that is so often associated with the theory-theory. If I am correct that the concept "knowledge" can be formed by contrast with examples of ignorance, for instance—and the concept of "belief" can then be derived somehow from "knowledge"—then there is at least one mentalistic concept, "knowledge," which is not functionalistic, even if others (like "belief") are.

Conclusion: Implications for deflationary naturalized epistemology

The possibility that children might grasp concepts of intentional mental states by first grasping *epistemic* states has not been widely entertained by psychologists or philosophers, probably because of the longstanding tradition in philosophy of thinking of knowledge as requiring analysis in terms of (justified true) belief, not vice-versa. But this tradition has recently been challenged by epistemologists like Timothy Williamson (2002), who argues that "know" should be understood as a primitive, whereas "believe" should be understood as putative knowledge or as a functional equivalent of knowledge.[35] If my account of folk psychology as originating in folk epistemology is correct, there is psychological support for Williamson's hypothesis. Simulation theorists will be disappointed, but theory theorists will also need to reconsider the widespread view that philosophy of mind is independent of epistemology.

Interestingly, in considering simulation theory as a naturalization of "belief," we have also now come full circle. Naturalized epistemologists begin with an interest in naturalizing knowledge. Because philosophic tradition says that knowledge is to be understood as a kind of belief, and because

[35] The present account of how the concept of "believe" develops from the concept "know" also helps enrich Williamson's claim that "know" has greater psychological explanatory value than "believe": if we need to understand what a subject knows and what he does not know in order to understand his beliefs, then knowledge attributions explain everything that belief attributions explain, and then some.

understanding belief has always posed an independent challenge to naturalists, naturalists then seek to naturalize belief. In chapter three, we explored various proposals for naturalizing belief in an object-naturalist style, but found them lacking. In the present chapter, we have now explored the most likely subject-naturalist proposal, simulation theory, and found it lacking, too—because there seems to be good reason to think that simulation itself cannot be performed or understood without the possession of epistemic concepts. In a way, this consideration calls the very project of naturalized epistemology into question. If the concept of "knowledge" is so basic as to ground even our understanding of "belief," perhaps it is not a concept that is in need of explanation, much less naturalistic explanation.

At this point, however, naturalists might reasonably ask whether knocking down simulation theory as a naturalization proposal is knocking down a straw man. Perhaps there are other subject-naturalist proposals out there worth examining, which may not depend on folk epistemology in the way that simulation theory seems to. This point is fair enough. In what remains of this section, I want to argue that even if there are other interesting descriptions of folk psychological practice that might be conducive to subject naturalization of belief, the facts we have presently uncovered about the apparent dependence of folk psychology on folk epistemology have some interesting implications for deflationary naturalism about *knowledge.*

In chapter 1, I examined the views of Michael Williams (1996), whose views I classified as a kind of pessimistic naturalized epistemologist. Although Williams does not characterize himself as a naturalized epistemology, he does take a deflationary approach to knowledge that would recommend itself to the subject naturalist, in the manner described by Price. Williams opposes what he calls "epistemological realism," the view that knowledge is a real thing common to all of the instances we call "knowledge." Instead he argues that the task of epistemology, if anything, is to examine our actual practices of knowledge attribution, for instance by examining how standards of justification shift from the context of one discipline to another. The case for this kind of deflationary naturalism about knowledge of course depends on the idea that knowledge is *not* a real thing (or that we have no reason to think that it is), and Williams supports this by claiming that cases of knowledge have no "theoretical integrity." What I will urge at present is that, interestingly enough, the evidence we have unearthed to oppose the deflationary view of "belief" helps to undermine Williams's deflationary view about knowledge by showing how knowledge *can* have theoretical integrity, though in a different way than Williams imagined it might.

Williams supports his contention that "knowledge" has no theoretical integrity by giving the example of Francis Bacon's early account of heat, which begins by listing a number of examples of heating: heating by radiation, by friction, by exothermic reactions, and by hot spices on the tongue. This list, says Williams, is at best a nominal kind. From the fact that we call all of them "hot" does not mean they all possess the property of *heat*. Likewise, simply because we say we know a number of things does not mean that there must be a theory of all things called "knowledge" (1996, 106–7). Of course in the case of heat, we know that there does end up being *some* coherent natural kind, described by the kinetic theory of heat: it just doesn't necessarily correspond to our pre-theoretic intuitions about "heat." So Williams goes deeper still, to argue that simply because a concept is teachable does not imply that it refers to something real. He mentions that the distinction between analytic and synthetic is teachable (a point made famous by Grice and Strawson in their critique of Quine), but that this does not imply that there really is a distinction between sentences called "analytic" and ones called "synthetic." For after all, the distinction between "witch" and "non-witch" was once teachable, but there are no witches (107–8). Presumably, (I am filling in some gaps in the argument here) Williams's point is that while scientists eventually found an underlying theoretical unity to *much* of our concept of "heat," there was none to be found for "witch." I take it that the application to "knowledge" is that, since scientists have not found anything equivalent to the kinetic theory of heat for "knowledge," the concept is more like "witch" than it is like "heat." Without any positive grounds for believing in theoretical integrity of the concept, if all we have learned is the mere ability to use the concept to make various distinctions (like that between knowledge and ignorance) we have no reason to believe that "knowledge" refers to anything real.

But Williams considers an objection to this argument. Simply because we do not have a theory of knowledge anything like our theory of heat does not mean that knowledge does not exist as a real thing. Consider things like tables and chairs. We have merely "loose, functional classifications" of these everyday objects, nothing like a rigorous physical theory about them. Yet we would never say tables and chairs do not exist (109). Williams responds that this objection assumes that "knowledge of the external world" is more like "chair" than it is like "witch." What causes "witch" to fail to refer, he says, is its status as an "essentially theoretical" term. Essentially theoretical terms, says Williams, are ones which "we see no point in continuing to make, or even no way of drawing, once the theory behind them has been rejected" (109). Clearly what Williams has in mind here is very similar to the descriptivist theory of reference discussed in chapter 3, as exemplified in Lewis' view of the reference of theoretical terms.

Williams then needs only to argue that "knowledge" is essentially theoretical, like "witch" but unlike "chair" (and, presumably, that the theory behind it has no support). He claims it is essentially theoretical, because there is "no commonsense, pre-theoretical practice that this way of classifying beliefs rationalizes: its sole function is to make possible a certain form of theoretical inquiry, the assessment of knowledge of the world as such" (110). The concept that Williams says is essentially theoretical, however, is "knowledge of the external world," which he takes to be understood in the Cartesian sense in contrast with "experiential knowledge." This point recalls much of his previous development, which rejects Cartesian-style arguments from the "priority of experience" to conclusions about the external world, and with them, the possibility of foundationalism. The possibility of foundationalism, Williams has argued, is the one hope the concept of "knowledge" has for exhibiting "theoretical integrity": if all of our knowledge can be shown to reduce somehow to the senses, this would be an important fact that all of it has in common. So if we follow Williams in rejecting foundationalism, and if its validity is a commitment of the concept of "knowledge of the external world," then it seems we should indeed decide that the very concept of "knowledge of the external world" fails to refer to some real fact common to all things we call "knowledge of the external world."

It is important to note, however, that this position assumes that the only relevant concept of knowledge is the Cartesian concept, and that the only relevant kind of foundationalism is the sort based on the "priority of experience." Arguably, there are other versions of foundationalism available which may not face the same problems as traditional Cartesian foundationalism. In particular, I have in mind direct realist foundationalism, which holds that basic beliefs are not beliefs about experience, but beliefs about ordinary middle-sized objects. (In general, I think many of Williams's objections to foundationalism trade on a confusion between versions which claim that *beliefs are based on experience* (which is definitive of any kind of empirical foundationalism) and versions which claim that *beliefs are based on <u>beliefs about</u> experience* (which is specifically Cartesian foundationalism). Now in fact I offer some arguments for direct realist foundationalism in my final chapter, but I do not need them presently to make the following point. As long as there are distinguishable versions of foundationalism, the failures of Cartesianism do not imply the failures of every foundationalism, and there is hope for the theoretical integrity of "knowledge" yet.

I can take this argument one step further. Even if many philosophers have assumed that the concept "knowledge" has Cartesian implications, this does not mean that the concept itself is committed to these implications. Just because the concept is not as non-theoretical as "chair" does not mean it is as theoretical as "witch." After all, what about "heat"? Williams himself seems to think that "heat" is something of a middle case between essentially theoretical and non-theoretical. Even though it was originally taken to have implications that we eventually rejected (e.g., that heat was caused by caloric fluid), we did not conclude that there is no heat. Williams calls concepts like these "classifications that have been theoretically rationalized but which retain independent utility" (110). "Distinctions like this," he says, "are apt to survive the rejection of theories with which they have become associated" (110) Williams does not explain what makes this middle case possible; perhaps it happens when a theory's core commitments are retained while others are rejected.

Whatever is to be made of this possibility, Williams at least thinks highly of it, and offers to show that it does not apply to "knowledge." He says that the concept has no "pre-theoretical utility" or a "theory-independent way of drawing even approximately the right boundaries around it." But here I think Williams is just wrong, and the evidence we have considered about the dependence of folk psychology on folk epistemology proves it. That evidence suggests that at an age when children are far too young to have read Descartes, or even to consider the idea that they are only aware of their internal experience, not the outer world, they are still "little foundationalists." They come to see perception as the source of knowledge, and cite it as explaining how they know various claims. Not only that, but we have also now seen that being able to attribute knowledge or the lack of it is a prerequisite of attributing beliefs, and insofar as attributing beliefs has predictive and explanatory power, then so too does attributing knowledge or lack of it. Finally, the suggestion that there is not even an approximate way of drawing the boundaries of the concept knowledge is just false. Perhaps it is difficult to draw the boundary between "knowledge of the external world" and "experiential knowledge" (where the latter is taken to mean knowledge *about* sensory experience), but this presupposes an unnecessary (and unworkable) Cartesian foundationalism. If children start out not with beliefs about that contrast, but simply about the contrast between knowledge (full stop) and *ignorance*, that is a perfectly acceptable basis for drawing the boundaries of the concept, even if it does not tell them everything they need to know right away.

So what I am suggesting is that even if philosophers' theories of knowledge have had unfortunate implications in the past, studying actual folk epistemological practices (as the subject naturalist tells us to do!) reveals that these implications do not exhaust the theoretical integrity of the folk concept of "knowledge." For that reason, it is really much more like "heat" than it is like "witch." As a result, there is more reason to think that difficulties in philosophers' theories of knowledge make no difference to the fact that there is knowledge. As a consequence, Williams's case against epistemological realism is undermined, and the motivation for his deflationary naturalism about "knowledge" loses its motivation. This is why undermining the version of deflationary naturalism about belief that we have discussed is also relevant to undermining deflationary naturalism about knowledge—even if there are other possible versions of belief deflationism available that I have not yet considered.

Now that the first form of pessimistic naturalism has been called into question—and given that optimistic naturalisms have been discredited since chapter 2 and 3—we have no choice but to turn to the final version of naturalized epistemology: Quinean pessimistic naturalism. In the next chapter, I will examine what distinguishes this version of naturalism from others, by describing its aims and its roots. In the final chapter, I will argue that its roots grow from both scientific and philosophical errors, and that if we can conceive alternatives to them, we need not resort to Quinean pessimism—or any version of naturalism, for that matter.

Appendix: Gordon on reason explanations and counterfactuals

In the chapter above, I think I have called into question the possibility of a simulation-theoretic explanation of the predictive and explanatory power of folk psychology. There is, however, recent work by Gordon that might be taken to establish that simulation has greater explanatory resources than I might originally have thought. In more recent work, Gordon has put forth a view of simulation theory that might circumvent some of the problems I have raised above. In "Simulation and Reason Explanation: The Radical View" (2001), Gordon contends that radical simulation can provide important folk psychological explanatory power, provided that we retool our understanding of *explanation*. Perhaps, even if simulation does not have the *same* explanatory power as false belief attribution, it still has *some* explanatory power, and its function in our mental economy is worth considering after all.

Part of Gordon's project in "Simulation and Reason Explanation" is to defend his older view of simulation from a different objection: "explanations" produced by simulation appeal to *reasons for action*, not *causes*. Davidson (1963), however, has argued that there is an important difference between the first and the second, and that only the appeal to the causes can amount to an explanation. The presence of the grizzly bear in Tom's path may present a *reason* for him to run, a *justification* for his quick departure, but its presence does not necessarily explain his action. This last point is particularly salient for cases of overdetermination. When there is more than one reason for action present in a particular case, it is impossible to appeal to only one as the explanatory factor. It seems, therefore, that the only way to explain an action is nomological, to subsume it under some law of nature. To explain human action, the usual strategy is to subsume behavior under laws relating mental states and actions—not laws relating objects in the environment (like *bears*) and actions.

Of course the deductive-nomological view of explanation has suffered from numerous philosophical problems in recent years (see Salmon (1998)). Cognizant of this and of his need to find an alternative account of psychological explanation that overcomes the overdetermination problem, Gordon considers *counterfactual* explanation. An example of an overdetermined action would be Gordon's braking at an intersection. One reason to brake is that there is a red light; another is that he is driving ten miles over the limit. Both provide a good reason to brake, but which reason explains braking? Gordon says that it *is* the case that if there hadn't been a red light, he would not have braked. But it is not the case that if he hadn't been traveling over the speed limit, he would not have braked (for even if he hadn't been speeding, he might have had a red light). So appealing to the counterfactual differences between competing reasons for action does seem to dissolve the overdetermination problem, and explain actions.

Gordon says that this kind of counterfactual explanation is easily adopted in the context of simulation, for the purpose of explaining human behavior. We can see that already in the example concerning the explanation for Gordon's braking. Suppose that Sam wants to explain Gordon's action.

Sam can imagine himself in Gordon's situation, and begin to imagine counterfactual possibilities. In one iteration, he imagines himself driving in Gordon's car, over the speed limit and through a green light at the intersection. He does not brake. Then he imagines himself driving the speed limit, only this time the light is red. He does brake. Because of that counterfactual difference, in the context of his simulation, he is able to say that he brakes *because of the red light*, not because of driving over the speed limit.

Of course Gordon (2001) acknowledges that this kind of counterfactual explanation does not account for the kinds of behavior that false belief explanation might account for:

> Where the explanans . . . is a reason-in-favor, then for the explanation to be correct, I had to have known or been aware that there was smoke; therefore, I had to have believed there was. … In interpreting the counterfactual that corresponds to a reason explanation, we consider only worlds in which the counterfactual condition c, specified by the antecedent of the conditional, lies within the agent's epistemic horizon: the agent knows or is aware that c—and, therefore, believes that c. We don't allow the counterfactual 'facts' to vary independently of the agent's beliefs.

So it is clear that this counterfactual reason explanation adds nothing to overcome the "problem of adjustment" I have discussed in the body of the paper above. It does not furnish any new resources needed for turning simulation into full-bore false belief explanation. (Of course I have argued that even the simplest versions of simulation run up against the epistemological problem of adjustment, but I will leave that aside for the moment.) However, we are presently concerned with the question of whether counterfactual reason explanation can account for the explanatory value of *any* degree of simulation. If it could, then perhaps there is something to simulation theory after all. Perhaps it could provide for some simple explanations from which theories could be built, and the remainder of folk psychological explanation could be account for by theory-theory.

But once we remember that counterfactual explanation is supposed to provide a simpler type of explanation than false belief explanation, we are faced with a new problem. That is because, when we look at the literature in developmental psychology, it appears that young children have systematic difficulties reasoning with counterfactual conditionals. Indeed the ability to deal with counterfactuals emerges at about the same time as the ability to pass the false belief test (Riggs et al. 1998). This has led numerous psychologists to believe that the abilities are related, though there is much debate about the particular nature of the relation (see Mitchell and Riggs (2000)).

Therefore, even if counterfactual reason explanation is efficacious as a form of simulation, it is unlikely that it accomplishes this without whatever resources are needed for false belief explanation. Since we already have reason to think that genuine, radical simulation cannot account for false belief explanation, it is also therefore likely that counterfactual reason explanation is a form of radical simulation. More likely, it involves some kinds of doxastic presuppositions that put it on par with theory-theoretic explanation.

CHAPTER 5

QUINE'S ACQUIESCENCE IN SKEPTICISM

Quine's (1969a) arguments for naturalizing epistemology have sparked dissent from the ranks of traditional epistemologists. For example, Kim (1988) complains that by making epistemology a "chapter of psychology," the naturalist robs epistemology of its normative force and thereby its status of genuine epistemology. I agree with the spirit of Kim's objection, but it is far from clear that Quine or other naturalists would agree that naturalized epistemology must be merely descriptive rather than normative. Quine (1986a; 1992) denies it, and proposals for naturalizing normativity abound. Other critics like Stroud (1981; 1984) sympathize with the idea that Quine has simply changed the subject, if not by dropping normativity, then for other reasons.

Yet as early as "Epistemology Naturalized," Quine insists that naturalism still yields epistemology "or something like it," on the grounds that the naturalism studies a natural phenomenon which is the subject matter for traditional epistemological questions:

> [E]pistemology still goes on, though in a new setting and a clarified status. Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input—certain patterns of irradiation in assorted frequencies, for instance—and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a *relation that we are prompted to study for somewhat the same reasons that always prompted epistemology*; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence. (1969a, 82–3, emphasis mine)

Even though Quine thinks naturalized epistemology is still very much like traditional epistemology, he does stress its "new setting" and "clarified status." In chapter 2, I explained what Quine meant by each of these. The "new setting" is psychology, which implies, for example, that important questions about the relationship between evidence and theory are no longer to be settled by relating evidence and "awareness," but instead by relating evidence to causal proximity to sensory stimulation. The "clarified status" of epistemology is that it is no longer to be concerned with discovering or with deriving on its own any first principles, given that theory “transcends any available evidence” (a statement of Quine’s underdetermination thesis).

But this understanding of Quine’s naturalized epistemology is largely negative. It does not concern awareness and it does not concern first principles. So what, then, does Quine want to achieve through such a study? What kind of relationship between evidence and theory does he mean to examine? He admits that it would be circular to try to try to validate the grounds of empirical science by using science (Quine 1969a, 75–6). But he also claims that we should not *want* to try, since philosophers stopped dreaming of a deductivist foundationalism for science long ago (76). If the evidence-theory relationship to be studied is not the traditional relationship of epistemic justification, what then is it?

The critics suggest that whatever the project of the naturalized epistemologist, if it is to count as epistemology, it must at least confront the problem of skepticism in some manner. Critics might say that if Quine doesn’t intend to establish the foundations of science, then he should at least show how to “dissolve” the problem of skepticism, perhaps by offering a “therapeutic” diagnosis of skepticism in the manner of Wittgenstein (1969), by asking “whether it can make sense to doubt” what the skeptic asks us to doubt (2e). Yet Quine is ambiguous about his estimate of the Wittgensteinian strategy, pejoratively characterizing it as offering philosophers a mere “residual philosophical vocation.” He urges that after the death of foundationalism, contrary to this strategy, “epistemology still goes on” (Quine 1969a, 82).

Quine does appear to offer a naturalistic strategy for dealing with the skeptic, but the nature of this strategy is somewhat unclear. In “The Nature of Natural Knowledge” (2004d), he argues that skeptical doubt is indeed what prompts epistemology—but that skeptical doubt is itself a product of science.[36] He notes, for example, that illusions can only be identified as such in relation to the existence of “genuine bodies with which to contrast them;” likewise the attempt to account for

[36] See also Quine (1960, 2) and Quine (1974, 1-4).

awareness of a third dimension based on two-dimensional images on the retina could only be initiated against the backdrop of the investigation of three-dimensional physiology. But in observing this, Quine does *not* appear to be raising therapeutic points designed to show that the skeptic's doubts make no sense. He does not wish to say that the skeptic presupposes knowledge of the external world, and therefore is engaging in a self-defeating argument. Indeed, he notes:

> [The skeptic] is quite within his rights in assuming science in order to refute science; this, if carried out, would be a straightforward argument by *reductio ad absurdum*. I am only making the point that skeptical doubts are scientific doubts. (2004d, 288)

If Quine thinks that the skeptic is within his logical rights to assume science to refute science, this implies at least that there is nothing straightforwardly incoherent in such assumptions. There may, of course, be something incoherent drawn out of the assumption—which is why it would be a *reductio ad absurdum*—but this incoherency would be grounds for skepticism, not a refutation of it.

What then is the significance Quine ascribes to the fact that skeptical doubts are scientific ones, if it is not therapeutic, and how exactly is it supposed to permit him to answer the skeptic in such a way as to retain some remnant of the subject matter of traditional epistemology? Most importantly, can this strategy succeed? Critics think that Quine either has no logical right to the free use of science to answer the skeptic's challenge (on the grounds that such use would beg the question) or that on Quine's own terms, such use can only lead to skepticism itself. Defenders think Quine's strategy does not beg the question, and that it can succeed against the skeptic, in effect by dissolving skeptical worries after diagnosing their source.

In this chapter, I will examine the major criticisms leveled at Quine's strategy, the attempt of Quine and his defenders to reply to these criticisms, and then evaluate the replies. I will argue that in the end, many interpreters of Quine—including both critics and defenders—are confused about what Quine is trying to do in response to the skeptic. Critics assume his views lead to a skepticism he does not desire; his defenders assume his views can help refute the skeptic. Both parties, I will argue, are incorrect. Quine's views do lead to a version of skepticism, at least by the standards of traditional epistemology. But since Quine is not a traditional epistemologist, this will not worry him. By his own *pragmatic* standards, he is not a skeptic. The question is whether his own pragmatic standards are consistent with his naturalism, whether a pragmatic account of justification can privilege scientific discourse over all other kinds. I will give reasons to doubt this. If we are upset, then, about Quine's complete abandonment of the traditional goals of epistemology, and challenged by the difficulties of his pragmatism, we should then wonder if we need to be naturalists after all. In my final section, I will explore the roots in Quine's skepticism (traditionally conceived), and show how his various negative theses (including both the indeterminacy of translation and the inscrutability of reference) derive from an argument also used by his underdetermination thesis.

Quinean skepticism via underdetermination and inscrutability?

Barry Stroud was well aware of Quine's strategy to dissolve skeptical problems by making free use of science; still he found this strategy unpersuasive. In "The Significance of Naturalized Epistemology" (1981), Stroud worries that *what* is revealed by a naturalistic examination of the evidence-theory relationship is not conducive to answering the skeptic. For example, given the "meager input" of the sensory surfaces, Quine tells us that the output of belief in physical objects is a "posit," or as he had put it in "Two Dogmas of Empiricism," "comparable, epistemologically, to the gods of Homer" (1953b, 44). Quine of course believes in physical objects and not the gods of Homer, and says everyone ought to agree with him. But the origin of the physical object hypothesis is "shrouded in prehistory" (Quine 1960, 22), and although it has no doubt proved successful, Stroud wonders why we should take it for granted in the face of the skeptic.

Stroud notes that Quine, like G.E. Moore, is willing to assert that there are physical objects. The question is whether this assertion should be taken as contradicting the skeptic in any way. Of course Stroud thinks that Quine wants to explain *how* we know about physical objects from within the scope of science, by explaining the route from meager input to torrential output. But Stroud reminds us that to *explain* the origin of some subject's knowledge, two conditions must hold: we, the explainer, must know that the subject's belief is true; and we must be able to show that it is not an accident, that the subject's posit turns out to be true because of some connection to the truth. Yet Quine's naturalistic investigation is also supposed to reveal that our subject's sensory inputs are "meager" in comparison to his outputs. If our position is similar to the subject's—a point Quine emphasizes—then there is a serious question about whether we, the investigators, are therefore even in a position to fulfill the two conditions of explaining the origin of knowledge, not only for our subject, but for ourselves. Even our understanding of sensory inputs as meager—a scientific discovery—would end up being a posit of its own, one rivaled by alternate hypotheses, as well.

Stroud acknowledges Quine's (1974; 1975b) points about how skeptical doubts arise in a scientific context. But Stroud is then at a loss to see the advantage behind naming the scientific nature of skeptical doubts. It might be thought that by pointing out how an understanding of "illusion" depends on a prior scientific grasp of "reality," Quine intends to show that the skeptic is asking an incoherent question when asking how we know our perception is not merely an illusion. But Stroud points out that Quine's acceptance of the legitimacy of the skeptic's use of science for *reductio* discounts this possibility. If skeptical questions were incoherent, then scientific assumptions could not be relied upon for the sake of *reductio*. If Quine concedes this, Stroud observes that it is hard to make sense of what *further* use science could be put in answering this *reductio*. If, by assuming certain facts about sensory input (for example), we are led to some general skeptical conclusion casting all of our knowledge into doubt, we have already reached the conclusion of our *reductio*, and at that point it would seem the epistemologist is no longer within his rights to make free use of science.

Stroud acknowledges that perhaps Quine means to do something else to answer the skeptic. Quine's (1981a, 475) answer to Stroud offers a glimmer of an alternative proposal. His direct answer to Stroud's query about *how* he intends to answer the skeptic is presented as follows, by reframing his attitude towards the skeptic's *reductio*:

> Thus, in keeping with my naturalism, I am reasoning within the overall scientific system rather than somehow above or beyond it. The same applies to my statement,

> quoted by Stroud, that "I am not accusing the sceptic of begging the question; he is quite within his rights in assuming science in order to refute science." The skeptic repudiates science because it is vulnerable to illusion on its own showing; and my only criticism of the skeptic is that he is overreacting.

There is a point to this response, because there is an important difference between relying on science *after* a *reductio* of science has gone through (which would be unjustified), and appealing to science *within the scope* of the alleged *reductio*, in order to show that it simply does not go through to begin with. If, for example, the naturalist examines more science than the skeptic does, and concludes that because of facts about illusions unappreciated by the skeptic, their existence casts no doubt on our knowledge at all, this would be an apparently naturalistic means of blocking the conclusion of the skeptic's *reductio*.

Another critic of Quine, Michael Williams (1996), appreciates this point. Williams points out that by pointing out the overreaction of the skeptic, Quine could simply be making a traditional refutation of the argument from error, by noting, for instance, that just because our senses sometimes "deceive" us does not imply that we should never trust them (a point made by Descartes himself). But Williams is concerned that if this is all Quine means by pointing out that the skeptic is "overreacting," then it does not accomplish much. Arguments from illusion and error are not thought to be serious grounds for *radical* skepticism, anyway, so it does the naturalist little good to diagnose the fallacies behind them. If, therefore, there are other scientific grounds for doubt that lead to radical skepticism, these grounds must be something other than the argument from illusion.

Williams suggests that what Quine must have in mind is the additional point, mentioned after the discussion of illusions in "The Nature of Natural Knowledge" (2004d, 288) that "science tells us that our only source of information about the external world is through the impact of light rays and molecules upon our sensory surfaces," which is supposed to make us wonder how we arrive at torrential scientific output given only such meager input. But here Williams appears to side with Stroud: if *this* is the scientific assumption leading to skeptical doubt, then there is no way for the naturalist to block its skeptical consequences. Williams thinks that this is part of Quine's basis for accepting the underdetermination thesis—a thesis Quine *affirms*, rather than dissolving by claiming that it is an "overreaction." If this thesis suggests that our theories are not justified, then this leads to radical skepticism. Indeed not only does Quine hold to the underdetermination thesis, but making sense of it even appears to be one of the *purposes* of naturalizing epistemology: Quine says we examine the relationship between evidence and theory *"in order* to see how evidence relates to theory, and *in what ways one's theory of nature transcends any available evidence*" (1969a, 83, emphasis mine).

There are at least two major ways in which Quine's naturalism highlights sensory inputs as "meager" in comparison to our theoretical outputs, and seems to imply a more radical form of skepticism, as a result. The first is its embrace of the inscrutability of reference thesis; the second its embrace of the underdetermination of theory by evidence. Although Quine discusses mainly the inscrutability thesis in his "Reply to Stroud" (1981a), Stroud actually invokes a hybrid of the two in attempting to show that Quine cannot respond successfully to the skeptic on his premises (Stroud 1981, 465).[37] In the remainder of this section, I will briefly outline these two theses, and explain why

[37] The distinction between underdetermination and inscrutability is, in fact, sometimes ambiguous. When Quine discusses underdetermination in his (1953b), he slides almost imperceptibly into saying that beliefs in physical objects are also underdetermined by the evidence (this is his infamous comparison between physical objects and the gods of Homer). But of course the sentences in which we register our belief in physical objects are typically observation sentences. Observation sentences are the medium in which all observational evidence is expressed, and thus there is a puzzle about what it could mean to say that belief in physical objects could be underdetermined by evidence, when rival theories said to be underdetermined are said to be such in relation to evidence that is held constant—evidence held in the form of observation sentences. This puzzle has led some, such as Williams (1996) to declare that Quine's views here are simply incoherent. I think, however, that the puzzle can be reconciled with a fuller understanding of Quine's views about inscrutability. It is true that observation sentences are held constant in order to say that non-observation sentences are underdetermined. But from that perspective observation sentences are taken holophrastically. Quine later concedes that when observation sentences are not treated holophrastically, they are theory-laden (Quine 1992, 7-9). It is true that few of us outside of philosophy ever form theories about the existence of physical objects. This is why Quine thinks that a belief in physical objects is buried deep in history. When children learn the individuating apparatus of reference, they are in effect *inheriting* this theory from history. So it is true that observation sentences are held fixed in relation to underdetermined theory, but as such they are merely holophrastic and do not represent an actual belief in physical objects. Once one approaches them from an adult perspective, from the perspective of dividing the reference of onetime holophrastic sentences, objects "go theoretical," and are subject to similar underdetermination concerns. There is, therefore, a way in which the underdetermination thesis is *fundamental* to the inscrutability thesis. If the latter has skeptical implications, it is probably in virtue of the former.

they each seem to imply a form of skepticism. Once this is complete, we will be in a position to evaluate Quine's full reply to skepticism in the next section.

The inscrutability of reference thesis is actually discussed explicitly in an exchange between Stroud and Quine. Stroud is worried that on Quine's view, there is a "possibility that the world is completely different in general from the way our sensory impacts and our internal makeup lead us to think of it." In discussion with Stroud, Quine proposes to understand this point in terms of "proxy functions and displaced ontologies" (1981a). "Proxy functions" are particular logical devices Quine has exploited to bolster the argument for his inscrutability of reference thesis, the thesis that the reference of individual terms within sentences are indeterminate, i.e. unsettled by the totality of evidence or sensory stimulation—the only relevant "facts of the matter"—available to language users (Quine 1969b). Unlike the indeterminacy of translation thesis, Quine takes inscrutability of reference to apply not only to terms in theoretical sentences, but those in observation sentences, as well. Famously, "rabbit" might refer to "rabbit," but also to rivals such as "rabbit stage," or "undetached rabbit part," or "Rabbithood" (Quine 1960). Quine uses proxy functions to show how similar rivals may be constructed for any term, simply by systematically mapping each predicate in a language to a unique predicate in a rival language.

Inscrutability *seems* to have inescapable skeptical consequences. If it is true that our terms may refer to any of Quine's proxies, then we don't know what the objects of our theoretical or even our observational sentences are, and we would seem to be cut off from the world. The reader may be concerned that inscrutability of reference by itself is not sufficient to generate serious skepticism. So what if we don't know *what* we're talking about?: perhaps all that counts is what we know about *whatever* it is we're talking about. But questions of reference and questions of epistemic justification might not be isolated so easily. It is likely that, at least in traditional epistemology, self-consciously successful reference is critical in avoiding Gettier problems. One way that a justified true belief can be only accidentally true is if it does not refer to the fact in question that makes the belief true. Suppose, for example, that Jones sees a look-alike of Smith in the room, and claims that Smith is in the room, so fails to know this even though Smith is in the room (he's hiding). Here we can diagnose the failure as resulting from the fact that the person to whom Jones is referring is not actually Smith.

Just in case there is still doubt as to the role of inscrutability of reference in generating skeptical doubts, there is of course another issue in Quine's philosophy that appears to have similar skeptical implications: the underdetermination thesis. This thesis claims, roughly, that there are very different scientific theories which are supported equally well by all available empirical evidence. As mentioned earlier, the underdetermination thesis is, of course, one of Quine's motivations for naturalizing epistemology in the first place.

Underdetermination's apparent skeptical implications have been explained best by Lars Bergstrom (1993, 344–5). Assuming that knowledge is justified true belief, Bergstrom argues that the existence of theories that are rival to, possibly incompatible with, but equally well-supported by the evidence as one's own theory, undermine one's justification in believing one's own theory—and thus, one's knowledge. This is particularly clear in a case in which there are two clearly theories, T_0 and T_r, which are equivalent in evidence and theoretical virtues, but known to be incompatible.[38] Knowing that T_r rivals our home theory T_0, Bergstrom says the only rational option is to suspend judgment between the two. We cannot justifiably pick one or the other, and so we do know which one is true.[39]

Bergstrom (2004, 105) has pointed out, however, that the skeptical implications of underdetermination are clearest only when the rival empirically equivalent theories are taken to be incompatible, i.e., not possibly both true. Earlier (1993, 343, 345), he does suggest that if T_0 and T_r and not incompatible but merely different, there might still be something irrational about accepting one rather than the other when both account equally well for the evidence. But it is worth pausing on this point. If T_0 and T_r which are empirically equivalent but not logically incompatible, it is trickier to

[38] The incompatibility enters, presumably, because of a difference in theoretical claims, e.g. about unobservables. After all, the underdetermination thesis is the underdetermination of *theory* by observational evidence.)
[39] Bergstrom thinks the same is true even if we don't know the nature of T_r, but simply know it exists (if, for example, we accept Quine's underdetermination thesis).

show how skepticism would arise, because it is then possible to take what Quine has called an "ecumenical" (rather than a "sectarian") line, and say that both of these theories could be true (Quine 1986b). If both can be said to be true, then it seems there is no question to be agnostic about, no reason to think one's present theory is threatened. In fact Quine believes that many cases of apparently incompatible rivals can be reduced to compatible ones, if incompatibilities arising from theoretical terms are eliminated by spelling the relevant theoretical terms differently in each theory (e.g., "the universe expands" vs. "the youniverse does not expand") rendering claims predicated with them compatible. Bergstrom gives a variety of reasons (related to simplicity and economy) for thinking it strange that one could be warranted in believing the conjunction of T_0 and T_r to be true (1993, 347), and argues that the spelling expedient would not *eliminate* the existence of incompatible theories—if there are such—but only allow us to deal with their compatible counterparts (350–1). In any case, Quine later (1986; 1992, 99–101) distances himself from the ecumenical position, though he leaves open the possibility that a sectarian might oscillate back and forth between compatible rivals, without believing both at the same time.

There are some philosophers, particularly those enamored of verificationism, who think there could never be such things as empirically equivalent but incompatible rivals (Dummett 1973, 617; Davidson 1990b, 306). The issues involved in deciding whether they are correct are difficult and beyond the scope of this essay.[40] I do think that if empirically equivalent rivals are never incompatible, this would make the skeptical consequences of underdetermination less obvious. And Roger Gibson has pointed out that empirically equivalent but compatible theories might still count as instances of underdetermination, according to Quine, because Quine (1988) also stresses that underdetermined theories are such that they cannot be rendered logically *equivalent* (120). Even if empirically equivalent rivals are not incompatible, this does not mean they can be rendered equivalent (as through the spelling expedient). If Gibson is right, then the underdetermination thesis would not be

[40] See Bergstrom (2000, 101-4) for a summary of the debate.

contradicted by the compatibility of empirically equivalent rivals (though note that this leaves untouched whatever genuinely incompatible theories there may be, prior to the spelling expedient).

At the same time, however, I think the version of the underdetermination thesis that does not assume the incompatibility of rival theses becomes trivial or at least uninteresting from an epistemological perspective. Here it is important to remember *why* the underdetermination thesis is so important in Quine's philosophy in the first place. As early as "Two Dogmas of Empiricism" (1953b), Quine invokes the metaphor of a field of force (theory) underdetermined by its boundary conditions (experience), in order to show that there is no such thing as a belief that may be held true come what may, or a belief that is immune from revision. In an earlier chapter, I have argued that the underdetermination thesis is also crucial to solidifying Quine's critique of traditional epistemology. It is an important element of his critique of the "doctrinal project" in epistemology, the attempt to show how one's knowledge can be justified by experience, insofar as it eliminates as options any number of foundationalist proposals, including merely probabilist (as opposed to Cartesian) candidates. If underdetermined (but compatible) theories are all equally true, Quine has many fewer ways to argue that the doctrinal project of traditional epistemology has failed. Likewise, as I shall argue later, underdetermination, or at least underdetermination-style theses, are also crucial in establishing the indeterminacy of translation, and thereby undermining the "conceptual project" in epistemology", which I have also argued is another central motivation for naturalizing epistemology. If underdetermined theories are all equally true—as the ecumenical position suggests—then there is no reason to say there is no fact of the matter involved in choosing a translation manual. If all translations are equally right, then there are many facts of the matter on which they are passed, and translation is not indeterminate. So, if underdetermination is watered down to not require the incompatibility of empirically equivalent rivals, it is true that its skeptical implications become less clear, but by that same token it also loses its significance as a motivation for naturalizing epistemology. We could be satisfied with undermining underdetermination this early in the game, but since I think there is much more of interest to explore, I will assume a version of the underdetermination thesis that motivates naturalized epistemology, we should examine the version that assumes that empirically equivalent rivals are incompatible. Indeed as late as "The Empirically Equivalent Systems of the World" (1975a) Quine does stress the *possibility* of incompatible rivals, even if not *all* empirically equivalent rivals are incompatible.

Between the inscrutability thesis and the underdetermination thesis, there does seem to be some support for Stroud's and William's contentions that Quine's resort to the scientific nature of skeptical doubt will do little to erase the skeptical implications of his own basic philosophic commitments. In the next section, however, I will explore responses offered on behalf of Quine himself and his supporters, which appear to undermine much of the force of the skepticism discussed above.

Quinean responses to skeptical challenges

Before endorsing Stroud or Williams, then, it is worth examining the fuller context of Quine's reply to Stroud, which might give us a better understanding of what Quine means when he says that the scientific nature of skeptical doubt shows that the skeptic is overreacting. Roger Gibson's (1988) defense of Quine against Stroud could shed some light here. Gibson maintains that Stroud's (1984) criticism of Quine (a recapitulation of much in Stroud's earlier article (1981)), fails because it neglects a central aspect of Quine's naturalism, the "reciprocal containment" of ontology (natural science) in epistemology and epistemology in ontology (natural science) (Quine 1969a, 83). The first containment is the view shared by both naturalized and traditional epistemology: the idea that we formulate our ontologies based on our accepted methods of acquiring knowledge. Just as the traditional epistemologist sought to construct science out of sense data, confining himself to its ontological deliverance, so the naturalized epistemologist acquiesces in the deliverances of the best science, because he accepts that knowledge only arises from the senses. But the second containment, of epistemology in ontology, is distinctive to the naturalist.

Gibson argues that Stroud is neglecting the significance of this second containment, which implies that naturalized epistemology presupposes the existence of the external world, including the sensory inputs which it judges to be meager (59). Why does it matter that epistemology presupposes these claims? Gibson elaborates:

> The relevant point about the containment (of epistemology by ontology) is that transcendental epistemology is incoherent. The skeptic may indeed use a portion of science to bring doubt to bear upon science, but only by presupposing the truth of other portions of science. For example, the skeptic might show that some scientific posits are epistemologically unwarranted, but his epistemological deliverances presuppose his *interim* acceptance of other scientific posits, namely, those presupposed by his own theory of evidence (59–60, emphasis mine).

So, Stroud and others may be worried that the inscrutability thesis leads one to raise skeptical doubts about the ontology (by way of the reference) of one's beliefs. But Quine's response is that even in the act of doubting our ontology of rabbits, given the possibility of a "rabbit stage" ontology, we are still presupposing as fixed the ontology of nerve endings, etc., which leads us to see a disparity between meager input and torrential output in the first place. Quine makes precisely this point in "Things and Their Place in Theories" (1981b, 21) an essay which appears to have developed out of his original critique of Stroud:

> Epistemology, for me, or what comes nearest to it, is the study of how we animals can have contrived that very science, given just that sketchy neural input. It is this study that reveals that displacements of our ontology through proxy functions would have measured up to that neural input no less faithfully. To recognize this is not to repudiate the ontology in terms of which the recognition took place.

So on this view, even if we can doubt some things, we can't doubt everything all at once. Therefore even if inscrutability and underdetermination lead us to be skeptical about some things, *radical* skepticism is, indeed, an overreaction.

But, we might object, how stable is the science that is "presupposed" by skeptical doubts? At minimum, to "presuppose" means that we *used* to believe it, up until the point that it came to a general conclusion about the reference of all terms, *including* the scientific ones used to formulate the original argument. This kind of presupposition still works perfectly well as a premise in a *reductio ad absurdum*, and yet it is a premise we might eventually come to reject as a result of that *reduction.* Note, after all, that Gibson says that what is presupposed by naturalized epistemology is *interim* acceptance of scientific theory. Quine himself (1960, 4) admits this much in a passage Gibson quotes immediately after making his point about interim acceptance:

> [O]ur questioning of objects can coherently begin only in relation to a system of theory which is itself predicated on our interim acceptances of objects. We are limited in how we can start even if not in where we may end up.

If "where we may end up" is not limited in the way that we start, that would seem to include "ending up" abandoning the objects we originally accept. Indeed, the wider context of this quotation from *Word and Object* suggests that this is a possibility Quine had in mind. He says that while we all start, like Dr. Johnson, affirming the existence of physical objects, we may come to find that best account of the world does not affirm this. Immediately after the sentences quoted by Gibson, Quine tells us: "To vary Neurath's figure with Wittgensteins' we may kick away the ladder only after we have climbed it" (1960, 4). This is clearly allowing for the possibility of kicking that ladder away.

Now Quine might have something other than skepticism in mind here, perhaps instead the possibility of coming to see the world composed of particles instead of commonsense objects. But in what follows he explains more about what he thinks could account for kicking the ladder away. Two paragraphs later, Quine explains that by beginning with physical object talk, we are merely assimilating a "cultural fare," without distinguishing between actual stimuli and what is posited additionally over and above them. He concludes (1960, 5):

> Retrospectively we may distinguish the components of theory-building, as we distinguish the proteins and carbohydrates while subsisting on them. We cannot strip away the conceptual trappings sentence by sentence and leave a description of the object world; but we can investigate the world, and man as a part of it, and thus find out what cues he could have of what goes on around him. Subtracting his cues from his world view, we get man's net contribution as the difference. This difference marks the extent of man's conceptual sovereignty—the domain within which he can revise theory while saving the data.

Quine's reference here to "the domain within which he can revise theory while saving the data" is yet another reference to his underdetermination thesis, or to his inscrutability of reference thesis. The continuing relevance of this point to Quine suggests that the containment of epistemology by ontology has little force to prevent the kind of *reductio* based on the underdetermination thesis, which Stroud envisions. Thus it seems we should concur with Stroud, along with Davidson (1990a, 74) and Koppelberg (1998, 266–7) who urge that the containment of epistemology in ontology is no panacea for the naturalistic response to skepticism. If our ontology contains scientific facts that suggest that we might abandon existence claims about physical objects—including the object presupposed by that scientific theory—then it seems we can, in fact, "kick away the ladder." Inscrutability and underdetermination do presuppose interim acceptance of scientific theory, but every *reduction ad absurdum* presupposes interim acceptance of whatever is to be reduced to absurdity.

Of course Quine has insisted he is not trying to challenge the coherence of the skeptic's doubts, so it is still unclear what he thinks the overall import of this scientific presupposition is supposed to be.[41] So perhaps we should find some further interpretation of his claim to block the "overreaction" of the skeptic. Indeed it is odd that both Stroud and Gibson focus their respective interpretations on only the final three paragraphs of "Reply to Stroud," neglecting to discuss the main body of the essay. There is much, in fact, to suggest that even though Quine knows that we are free to "kick away the ladder," he does not think this implies any skeptical threat. Immediately after saying that accepting the possibility of proxy functions does not imply that we must repudiate our ontology, he also says:

> We *can* repudiate it. We are free to switch, without doing violence to any evidence….But it is a confusion to suppose that we can stand aloof and recognize all the alternative ontologies as true in their several ways, all the envisaged worlds as

[41] I myself am sympathetic to the idea that the skeptic's doubts are incoherent, and that this by itself is sufficient to diagnose the skeptical illness. But I also agree with Stroud (1984, 227) that Quine's "view of language and his rejection of the philosophical use of synonymy or analyticity leave him in no position to appeal to what is or is not included in the meaning of a particular term," and that arguments from coherence do tend to presuppose specific theories of meaning, whether analytic or otherwise.

> real. It is a confusion of truth with evidential support. Truth is immanent, and there is no higher. We must speak from within a theory, albeit any of various. (1981a, 21)

Earlier in the essay, Quine makes a point that he also made directly in response to Stroud. He tells us that even when observation sentences are no longer treated holophrastically, but instead as composed of referring terms—even after we become adult philosophers and catch a glimmer of the possibility of replacing our terms with proxies—there is a way in which we are insulated from the effects of inscrutability:

> The point is not that we ourselves are casting about in vain for mooring. Staying aboard our own language and not rocking the boat, we are borne smoothly along on it and all is well; 'rabbit' denotes rabbits, and there is no sense in asking 'Rabbits in what sense of "rabbit"?' Reference goes inscrutable if, rocking the boat, we contemplate a permutational mapping of our language onto itself, or if we undertake translation. (1981a, 20)

Here Quine appears to be pulling back from a brink reached in "Ontological Relativity" (1969b), in which the notion of reference—one originally thought to be respectable and objective in "Notes on a Theory of Reference" (1953a)—appears to lose all such respectability and drop to the status enjoyed by murkier notions such as meaning. But Quine had hinted at this retreat even in "Ontological Relativity" when, arguing that questions of the reference of terms are answered only by translating them into other language, the resulting regress of translations could be halted only by "acquiescing in our mother tongue and taking its words at face value" (1969c, 49). Apparently what we do to refrain from rocking the boat with concerns over inscrutability is simply to acquiesce in this manner. Once we do this, we are able to maintain a "robust realism" about the reference of our terms, and affirm an "unswerving belief in external things—people, nerve endings, sticks, stones" (1981b, 245). This Quine sees as a reflection of his naturalism, the idea that truth is "immanent" to theory, that "it is within science itself, and not in some prior philosophy, that reality is to be identifies and described" (1981b, 246). So even if we can permute our preferred reference scheme with proxy functions, Quine's point is that we need not do what we can. By acquiescing in our mother tongue, we in effect accept a scheme of reference, and there is simply no question about other possible schemes.

Presumably a similar story can be told about our acceptance of theories, *mutatis mutandis*, that would obviate worries concerning underdetermination.

But does this strategy of acquiescence, based on the difference between what we can do and what we in fact do, provide a response to the skeptic that would satisfy the traditional epistemology? Retracing the steps by which Quine first formulated his inscrutability thesis suggest it is not. Consider one of his earliest formulations in *Word and Object* (1960, 51–2):

> For, consider "gavagai". Who knows but what the objects to which this term applies are not rabbits after all, but mere stages, or brief temporal segments, of rabbits? In either event the stimulus situations that prompt assent to "Gavagai" would be the same as for "Rabbit". Or perhaps the objects to which "gavagai" applies are all and sundry undetached parts of rabbits; again the stimulus meaning would register no difference.

The argument here is roughly parallel to Quine's argument for the indeterminacy of translation of whole sentences (except of course that Quine thinks inscrutability applies to terms even in observation sentences, while indeterminacy does *not* apply to observation sentences taken as wholes): he implies that we do *not* know the reference of the term "gavagai" (and later, "rabbit" itself) because the term could be equally true of rabbits, rabbit stages, etc., *given the same stimuli.* In other words, Quine wants to say there is no "fact of the matter" to determine reference, given that the only naturalistically respectable facts to consider are sensory stimuli and dispositions to assent, and these stimulations are logically compatible with any number of possible reference schemes.[42] This is parallel to his argument

---

[42] Now it might be objected that the argument for inscrutability listed above, regarding the compatibility of different reference schemes with identical stimuli and speech dispositions, is not Quine's only argument; other arguments may not carry with them this kind of significance for what we can do as opposed to what we do in fact do. Here one might appeal to the remarks of one commentator (Ben-Menahem 2005, 266) who says that Quine has two separate arguments for inscrutability: one the "informal" argument from an inability to extract "individuation schemes" from stimuli and speech dispositions, the other a "formal" argument exploiting certain logical properties of expressions. In the presentation above, I have not systematically separated these types of arguments. Arguments concerning rabbit vs. rabbit stage, and resulting questions like "Is this the same gavagai as that?" are examples of the first kind, which is more concerned with the situation of radical translation. Arguments concerning "proxy functions," are examples of the second, which are more concerned with inscrutability in the home language and even with our own utterances.

However I do not believe these arguments are fundamentally different: both turn on what we *can do* given the naturalistic facts of the matter in the same way. A review of the "formal" argument shows why. This argument is supposed to be independent of the "informal" argument, because even if we could settle on a determinate individuation scheme for some terms, it would at best settle questions of reference arising from *direct ostension*: from reference made through pointing to objects, etc. Even these determinate individuation for indeterminacy of translation, where he likewise argues that any number of translation manuals for theoretical sentences are equally acceptable given identical stimuli and speech dispositions: since these stimuli and dispositions are the only relevant facts of the matter, and they are compatible with multiple translation manuals, there is no fact of the matter to decide between competing translation manuals.

So, is the fact that there is a difference between whether we *can* and whether we *do* permute our terms into proxies relevant to stopping the infiltration of inscrutability? From the above argument for inscrutability, I do not see why the difference is relevant to addressing the concerns of the traditional epistemologist. The entire argument for inscrutability derives not from what we *do* in fact do, but merely from what we *can*: we *can* use a variety of proxies in the same manner as our original terms, without doing violence to our stimuli and speech dispositions. This becomes clearer when, after

---

schemes, however, would not settle questions of so-called "deferred ostension" (Quine 1969c, 40-41), reference made to objects other than objects pointed to, either through causal mediation or the relationship of instantiating a universal. In particular Quine discusses the attempt to use deferred ostension with abstract singular terms to refer to abstract objects. (I myself do not share Quine's (1960, 269) view that we have an ontological commitment to real abstract objects, and I find his arguments for this (Quine 1947) to be unpersuasive. This is, however, an issue beyond the scope of the present paper. Suffice it to say that on any account of universals, platonist or otherwise, it is clear that the reference of abstract terms cannot be settled through *mere* direct ostension. So the problem here is not unique to Quine's idiosyncratic views about ontological commitment.) Considering the example of a "thoughtful protosyntactician" who wishes to refer to the sentence types involved in his proof theory, Quine argues that to make reference to these abstract sentence types, he could map expressions onto sequences, which can in turn be mapped onto numbers (like Gödel numbers) (1969c, 41-2). Quine then considers the arithmetician, who could map numbers onto various set-theoretic constructs, any of which could be consistent with laws of arithmetic (43-44). The idea is that neither these laws, nor any amount of direct ostension towards the protosyntacticians expressions will settle the reference of the abstract *types* to which he wishes to refer. So: even if the "informal" argument does not succeed in rendering the reference of ostensive terms inscrutable, the "formal" argument is supposed to render inscrutable the reference of non-ostensive terms.

I think the above "formal" argument is no different in principle than the earlier argument about individuation. In fact it is not a uniquely formal *argument*, but simply an argument that is *about* formal properties (of mathematics and set theory). Whereas Quine's "informal" argument held that the reference of observation terms could not be settled by ostension, his "formal" argument simply says that even if direct ostension could settle anything, it couldn't settle the reference of abstract singular terms: do they refer to sequences, or numbers, or sets of sets? Like the earlier argument this argument simply identifies a range of reference schemes one *can* adopt given a fixed set of allegedly naturalistic facts: the only difference is that in the second argument the set of naturalistic facts has been charitably extended to include those related to direct ostension (a charity Quine is, in the end, reluctant to extend). Facts related to direct ostension do not settle the reference of abstract singular terms, nor do any related to laws of arithmetic. And in case it is objected that the arithmetic constraints are somehow independent of naturalistic concerns, it must be recalled that for Quine, mathematics, like logic, is of a piece with natural science, and continually tested through repeated applications in science (Quine 1995) showing how reference can only be specified by bringing in (equally inscrutable) questions about identity and diversity ("Is this the same gavagai as that?," etc.), Quine observes the following:

> Two pointings may be pointings to a numerically identical rabbit, to numerically distinct rabbit parts, and to numerically distinct rabbit stages; the inscrutability lies not in resemblance, but in the anatomy of sentences. We *could* equate a native expression with any of the disparate English terms 'rabbit', 'rabbit stage', 'undetached rabbit part', etc and still, by compensatorily juggling the translation of numerical identity and associated particles, preserve conformity to stimulus meanings and occasion sentences. (1960, 53–4, emphasis mine)

Quine's argument for inscrutability, I conclude, turns on what we can do, not what we do in fact do. That is because it is an argument about whether there is a *fact of the matter* constraining what we do. The fact that we can permute our terms into any number of proxies reflects the fact that doing so is not inconsistent with our sensory stimuli and speech dispositions, i.e., it reflects the only facts that are facts of the matter. Thus there *is no* fact of the matter picking out one reference scheme rather than another. That is what the inscrutability thesis means: that there is no such fact. We are free to pick whichever reference scheme we like, including one that does not, perhaps, *remind us* of this thesis (the one according to which "rabbit" refers to rabbits). Whether or not we choose to remind ourselves of our ignorance does not change the fact of our ignorance. Simply saying that "rabbits" refers to rabbits does not create a *fact* about reference.

Given the above, it appears that every argument Quine advances for the inscrutability of reference really does diminish the significance of the mother-tongue acquiescence strategy, at least from the perspective of *obviating the worries of the skeptic, to the satisfaction of the traditional epistemologist*. But of course Quine's worries are not necessarily those of the traditional epistemologist. Let me suggest that each of Quine's responses to the skeptic can make a limited amount of sense, provided that we stop trying to understand him as pursuing the goals of traditional epistemology. Of course insofar as naturalized epistemology is understood in contrast with traditional epistemology in the first place, it might seem that we should never have thought of him as pursuing these goals in the first place. But it would seem that way only if we neglect that Quine has long been claiming that there is some subject matter that is shared in common by traditional and naturalized epistemology. Indeed the naturalists I called "optimistic" thought that epistemology could pursue traditional goals (for example, explaining how our beliefs are justified) simply by adopting unconventional, naturalistic means. Chapters 2 and 3 called much of this into question, of course, and we are now seeing the full nature of the break between the optimist and the pessimist. The pessimistic naturalistic epistemology demurs even of achieving (many) traditional epistemological goals: of showing how beliefs in our ontologies can be justified, i.e., logically justified (deductively or inductively). What it shares as a common subject matter with traditional epistemology is not the goal of the logical justification of beliefs, but simply *some explanation or other* of our beliefs. As it happens, the kind of explanation Quinean naturalists seem to have in mind is also a form of justification, only not *logical* justification, but *pragmatic* justification. In the next section, I will explain how each of the elements of Quine's approach to the skeptic is imbued with this pragmatism.

Pragmatism and naturalism

First, let us consider the reciprocal containment point, in conjunction with Quine's contention that skeptical doubts presuppose the acceptance of ordinary scientific ontology. What exactly does Quine mean by "presuppose" here? Clearly in order to get to the point of accepting the inscrutability and underdetermination thesis, one needs to have *once* accepted various putative truths of science. My point in the section above is that this does not guarantee that one accepts them any more. (It is possible to "kick away the ladder.") I also mention that inscrutability and underdetermination each have a perfectly *general* scope, i.e., each concerns every scientific term or theory, including the ones used to describe meager sensory input. Because of this, it seems that it might even be incoherent to accept *both* the scientific ontology *and* the results of the inscrutability and underdetermination theses—and that this incoherence might serve as a *reductio* of the acceptance of scientific ontology. But as I have continually emphasized, Quine claims that there is no such incoherence. How might he explain the incoherence away?

I think that the answer is that Quine's "presupposition" thesis can only make sense in combination with his "acquiescence" thesis, his idea that we face no problem of inscrutability when we acquiesce in our mother tongue. The second thesis helps us to see that there is, in effect, a use/mention confusion in the suggestion that there is an incompatibility between acceptance of inscrutability and acceptance of scientific ontology. When a naturalized epistemologist argues for inscrutability of reference, he *uses* some scientific terms in order to come to a conclusion in which he *mentions* that "rabbit" and "rabbit stage" are reference schemes compatible with the meager sensory input. When the naturalized epistemologist uses those scientific terms, he is *himself* acquiescing in his mother scientific tongue. If he should then turn to zoology and begin to reason about *rabbits*, he once again acquiesces in his mother tongue. And this acquiescence is fully compatible with his embrace of the inscrutability of reference, for that embrace merely mentions "rabbit" and its rivals; it does not use them.

The natural question to raise at this point is: what if the scientist, recognizing the inscrutability of reference, suddenly decides to abandon the disquotational reference scheme, and affirm that "rabbit" refers to rabbit-stages (mapping his own language onto itself)? Quine would have to allow this possibility. Of course it does not immediately threaten the scientific ontology used to generate the inscrutability thesis, since that ontology presumably contained nerve endings, etc., rather than rabbits. Even so, Quine would have to allow that decisions to abandon disquotation are possible given the recognition of inscrutability. Why, then, does he seem unconcerned? Here, I think, is where his pragmatism enters. Why pick "rabbit" rather than "rabbit stages"? Because a disquotational reference scheme is simply easier to apply. Perhaps there will be cases in which we avoid disquotation—perhaps we commit malapropisms—but these will be exceptional cases, and will make sense only against the background of lots of other disquotationally-generated reference. Accepting this reference scheme because it is easier does not imply, of course, that we now have some logical justification in terms of reference-facts; it only means we have pragmatic justification.

There is another dimension of pragmatism to the reference scheme we accept. When deciding which terms to *use*, our decision is pragmatic in the sense that disquotation is easy. But when the naturalized epistemologist *mentions* the terms we use on the meta-level, and discusses what it is we talk about when we use them, he may be inclined to say the following:

> To say what objects someone is talking about is to *say no more* than how we propose to translate his terms into ours; we are free to vary the decision with a proxy function…. Structure is what matters to a theory, and not the choice of its objects. F.P. Ramsey urged this point fifty years ago, arguing along other lines, and in a vague way it had been a persistent theme also in Russells' *Analysis of Matter*. But Ramsey and Russell were talking only of what they called theoretical objects, as opposed to observable objects. I extend the doctrine to objects generally, for I see all objects as theoretical….The objects, or values of variables, serve merely as indices along the way, and we may permute or supplant them as we please as long as the sentence-to-sentence structure is preserved. The scientific system, ontology and all, is a conceptual bridge of our own making, linking sensory stimulation to sensory stimulation. (Quine 1981b, 20, emphasis mine)

Or he may even say this:

> What then does our overall scientific theory really *claim* regarding the world? *Only* that it is somehow so structured as to assure the sequences of stimulation that our theory gives us [sic] to expect. More concrete demands are indifferent to our scientific theory itself, what with the freedom of proxy functions. (Quine 1981a, 474, emphasis mine)[43]

When speaking as a naturalized epistemologist on the meta-level, then, Quine seems to describe an *almost* fully instrumentalist or pragmatist semantics. I say "almost" because he does allow that our theoretical terms might at least refer to how the world is "structured so as to assure sequences of stimulation that our theory gives us to expect." Mentioning that structuring might be taken to mean

[43] It has been brought to my attention that if Quine accepts this as a serious statement about the content of scientific theories, it may have the effect of truly trivializing his statement of the underdetermination thesis. It would imply that empirically equivalent theories are also logically equivalent, and therefore certainly logically compatible. Of course, this is only if the statement is interpreted in a purely phenomenalist manner, not in the structural realist manner. As I've said, there's a fine line between the two. In any case, I think that Quine did have a tendency to entertain more and more trivial versions of underdetermination as the years went by. All I can say is that the more trivial they become, the less motivated naturalized epistemology becomes. The trouble is that while Quine lost confidence in underdetermination, he kept confidence in the project of naturalizing epistemology. This is trouble because without the first, there may have been little motivation for the second.

that we refer to underlying essences which somehow order our sensations, as in a two-factor theory of reference. But what is important for Quine, who disavows the naturalistic respectability of natural kinds, is that any number of possible reference schemes can exhibit the same structure. There is a fine line, then, between the possibility of the world's exhibiting the same structure through many different ontologies, and *our experience* having the same structure, regardless of the world's ontology. When speaking on the meta-level, it is hard to see whether the naturalized epistemologist is committed to structural realism or simply to phenomenalism. In either case, the pragmatic element is all that matters: what matters to speaking of objects is the role the play in permitting us to explain and predict our "sequences of stimulation." Even if the content of our statements about objects is not exhausted by their pragmatic role, the inferential *significance* of our statements about objects is.

To many, the idea that claims about objects are not primarily claims about external, mind-independent objects is already to concede everything to the skeptic. (Lars Bergstrom (1993, 255) argues something along these lines.) But skepticism is just the position that our beliefs (about the external world, or anything else) are not *justified*. So the same epistemology may look skeptical to those who hold one standard of justification, but non-skeptical to others. Because of his inscrutability and underdetermination theses, Quine cannot accept that there is full *logical* (deductive or inductive) justification for belief in our preferred theories or ontology. So he accepts what looks like skepticism from the perspective of the traditional epistemologist (who demands logical justification for our beliefs to avoid skepticism). But what he accepts is not skepticism by reference to a pragmatic standard of justification.

Quine's commitment to pragmatism is especially evident in his discussion of theory choice. While he thinks that theory is *logically* underdetermined by evidence, he thinks choices among empirically equivalent rivals are ultimately based on pragmatic concerns. In his later works, he enumerates a list of "theoretical virtues" possessed by our preferred scientific theories: conservatism, generality, simplicity, refutability, and modesty (1992, 20). Elsewhere Quine (1992, 15) explicitly links theoretical virtues like simplicity and "minimum mutilation" with predictive power:

> [T]he ultimate objective is so to choose the revision as to maximize future success in prediction: future coverage of true observation categoricals. There is no recipe for this, but maximization of simplicity and minimization of mutilation are maxims by which science strives for vindication in future predictions.

How, then, is Quine responding to the skeptic if he concedes that our beliefs lack full logical justification—even if they are supplemented by pragmatic justification? The difference between his mention of the skeptic's "overreaction" in "Reply to Stroud" and "Things and their Place in Theories" is revealing. In the first, he simply says that his "only criticism of the skeptic is that he is overreacting" (1981a, 475). In the second, he says the skeptic is merely "overreacting when he repudiates science across the board" (1981b, 22). The first excerpt doesn't specify the respect in which Quine thinks the skeptic is overreacting. This has led some (including critics and defenders) to conclude that Quine thinks the skeptic's embrace of *skepticism* is the overreaction. The second excerpt clarifies that it is simply the skeptic's repudiation of science that is the problem. If Quine is a pragmatist, however, he can embrace both the skeptic's thesis that our beliefs are not fully logically justified *and* embrace the pragmatic power of science that a skeptic might be inclined to reject on this ground. He can embrace the latter because even if science does not deliver full logical justification of our beliefs, in a way that circumvents inscrutability and underdetermination, it does deliver all the pragmatic justification we need to predict and control our experiences.

Accepting Quine as both a pragmatist and skeptic (understood in the traditional sense) helps to explain a number of other interpretative quandaries about Quine's response to skepticism.

One way in which Quine responds to Stroud's worries about the skeptical implications of inscrutability is to find solace in the fact that truth of observation sentences is prior to reference on this thesis. Inscrutability allows that given our stock of observations, we may still assert a constant set of observation sentences, which we have no option but to regard as true. Even if we cannot divide the reference of our sentence in a determinate way, taken *holophrastically* they can still be asserted to be true. Truth here is understood in a deflationary manner: to assert a sentence as true is simply to assert it. We do not need correspondence relations to understand truth.

But at first this response seems unresponsive to Stroud's claim. Stroud has claimed that this is a straightforward endeavor to study the relationship between sensory input and theoretical output, provided that we are in a position to observe important facts about the subject's environment and its relationship to him. If we see that what a subject claims is true, and if his claim is a reliable one, we have no problem explaining the origin of his knowledge. Stroud observes, however, that the mere truth of the subject's beliefs is not sufficient to explain this knowledge, although it is necessary (1981, 461). If the subject claims there are bodies, and we see there are none, we know the subject does not know there are bodies. But even if the subject's output is true, the subject does not necessarily know if it is only accidentally true. If, for example, the subject claims there are bodies in front of him, and there are—but they are behind a screen past which he cannot see—then his output is only accidentally true and he does not know. Stroud goes even further, and claims that if we have no access to truth or reliability of the subject's beliefs, we cannot decide whether they know. But if we glean from our naturalistic study of the subject's meager inputs that theory is underdetermined by evidence, we will come to see our own theories as likewise underdetermined, and we cannot understand how *we* can ever come to know anything, including the claims needed for evaluating whether or not the subject knows (462–3).

In response to this, for Quine simply to assert that one might assert true observation sentences in response to some evidence *seems* to be missing the point. So what if the observation sentences are true—even if they come along with some sensory evidence? Gettier cases are plentiful, and show that in traditional views of knowledge, justified true belief is not sufficient for knowledge. Presumably this is what Stroud is emphasizing when he claims that it is not enough that a subject's belief be accidentally true. But Quine is only missing the point here if Stroud is not. Stroud appears to believe that Quine is trying to offer a *refutation* of the skeptic, some positive naturalistic case for how we can show that both we and the subject can have logically justified beliefs. But if Quine is not even trying to leverage naturalism to answer traditional epistemological questions, then Stroud is missing the point, and Quine is not. Quine (1981a, 474) seems to suggest as much in the following:

> Stroud finds difficulty in reconciling my naturalistic stance with my concern with how we gain our knowledge of the world. We may stimulate a psychological subject and compare his resulting beliefs with the facts as we know them; this much Stroud grants, but he demurs at our projecting ourselves into the subject's place, since we no longer have the independent facts to compare with. My answer is that this projection must be seen not transcendentally but as a routine matter of analogies and causal hypotheses. True, we must hedge the perhaps too stringent connotations of the verb "know"; but such is fallibilism.

This is the paragraph that directly precedes the paragraph in which Quine laments the skeptic's "overreaction." Seen in this context, it should be especially clear that whatever Quine's response is to the skeptic, it has little to do with traditional epistemological responses to skepticism. Quine's last line, about not being too stringent about the verb "know" is particularly revealing. It shows that he is not interested in holding onto an epistemology that explains the origin of knowledge in a manner faithful to the "traditional" concept of knowledge (some kind of logically justified true belief). Quine is, after all, not much interested in "traditional" concepts of anything. He is not interested in conceptual analysis as a method in philosophy. He is interesting only in *explication*: the process of taking some pre-existing concept and modifying it to make it useful for theoretical purposes. Quine would probably say that the traditional "justified true belief" concept of "know" serves no important purposes, hence it is safe to discard.

Another interpretive quandary resolved by treating Quine as a pragmatist is the paragraph that immediately follows his paragraph about the skeptic's overreaction:

> Experience might, tomorrow, take a turn that would justify the skeptic's doubts about external objects. Our success in predicting observations might fall of sharply, and concomitantly with this we might begin to be somewhat successful in basing predictions upon dreams or reveries. At that point we might reasonably doubt our theory of nature in even its broadest outlines. But our doubts would still be immanent, and of a piece with the scientific endeavor. (1981a, 475)

Commenting on this passage, Stroud (1984) considers that it may have something to do with Quine's lament of the skeptic's overreaction. Indeed it does, but not in the way that Stroud proposes. Stroud suggests that Quine thinks that only if the predictive power of science wanes should we take the skeptic seriously. Since its predictive power has not waned, we should therefore reject skepticism. On this view, skepticism would be a doctrine that is itself subject to confirmation. Stroud notes that this is not what the skeptic says: the skeptic does not take a position about some rival source of our beliefs (e.g., dreams rather than science): he only says none of our knowledge, whatever its content, is logically justified. As a result, Stroud thinks that Quine's alleged answer to skepticism is knocking down a straw man, and ineffective.

But in light of my reading of Quine as a pragmatist who is willing to concede that our beliefs are not fully logically justified, we can interpret him differently here. Roger Gibson, a defender of Quine, castigates Stroud for alleging that Quine takes skepticism to be subject to confirmation (1988, 59). And surely skepticism understood in the traditional way, as thesis that our beliefs are not fully logically justified, is not a thesis that we would treat as subject to confirmation. Gibson goes on to object that Stroud misses the importance of reciprocal containment, which we have already discussed. Curiously, however, Gibson offers no alternative explanation for what Quine actually means in the passage about the possibility that our predictions might some day lose their power. Now I think we can offer an explanation. Even if the traditional thesis of skepticism is not subject to confirmation, a pragmatic version of the thesis might be. That is to say, if skepticism is the idea that our beliefs are not justified, then if we admit pragmatic justification as one species of justification, whether or not skepticism is true will depend on whether or not our beliefs are pragmatically justified. And *that* is a thesis that we can imagine being subject to confirmation. Whether or not our beliefs are pragmatically justified depends on whether or not they have significant explanatory and predictive power. In the passage quoted above, Quine is saying that we trust scientific beliefs because they *do* allow us to predict observations. If someday they stop allowing this, then we would no longer regard them as (pragmatically) justified. If dreams instead turned out to yield the best predictions, we would regard them as justified instead.

It should come as no surprise that Quine is willing to embrace the thesis that our beliefs are not logically justified—even if he thinks they are pragmatically justified. Quine (1969a, 72), after all, famously announced in "Epistemology Naturalized" that the doctrinal and conceptual projects of epistemology had failed, and that "the Humean predicament is the human predicament." Furthermore, there is the tantalizing line at the end of his description of making epistemology a "chapter of psychology," in which he mentions that one of the *purposes* of naturalized epistemology is to discover how "one's theory of nature transcends any available evidence." This seems to build the problem of underdetermination—and any associated traditional skeptical theses—into the naturalist project from the beginning.

After all, if we accept that we can never understand our beliefs to be fully logically justified, there is a problem: if our beliefs have not been determined by evidence, what then *are* they determined by? What makes it possible for us to have scientific output that is "torrential" in comparison to our "meager" perceptual input? Quine's answer is contained in *Word and Object*, *The Roots of Reference*, and his other works in which he describes the variety of accidental, analogical, and otherwise non-logical devices by which such theory is formed. It is not intended to show how our knowledge is justified, but how our "knowledge" arises in a world where justification in the traditional sense is not an option.[44] It is not even intended to show us how our sensory evidence connects us cognitively to independent facts. As Quine wrote in "Epistemology Naturalized," "Awareness ceased to be demanded when we gave up trying to justify our knowledge of the external world by rational reconstruction" (1969a, 84). The variety of accidental, analogical, and other non-logical devices that

[44] Note that this does not imply that Quine has necessarily abandoned epistemology as a normative project. As I have already suggested, he still has the option of naturalizing normativity, of showing how these various theoretical developments have served some adaptive function for us. So they would count as "good" beliefs from the perspective of natural selection, perhaps. Even so, they would only be normative in this new sense, not in the traditional epistemological sense of justifying our beliefs.

do explain the origin of our theory are *pragmatic* devices. As Quine (1953b, 46) remarks in "Two Dogmas of Empiricism":

> In repudiating [the boundary between the analytic and the synthetic], I espouse a more through pragmatism. Each man is given a scientific heritage plus a continuing barrage of sensory stimulation; and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic.

Stroud, Williams and others have noted the statements about the Humean predicament, and have considered that Quine might not intend to *refute* the skeptic. But they are usually perplexed by Quine's statements, in *The Roots of Reference* and elsewhere, which analyze skeptical doubts as arising from science. These in combination with Quine's claims that the skeptic is overreacting make critics think Quine is simply inconsistent: sometimes he concedes the full force of skepticism, other times he wishes to answer it. But if my reading is correct, Quine is not inconsistent. His remarks about the scientific source of (local) skeptical doubts are simply aspects of his attempt to show that even if we accept that our beliefs are not fully logically justified, this is no cause for concern. The skeptic overreacts by ignoring the possibility of pragmatic justification, which in turn explains why we are able to acquiesce in our mother tongue and mother scientific theory. This is satisfactory if you share Quine's affinity to pragmatism, but not otherwise. Since his acquiescence strategy does nothing to calm traditional skeptical doubts or show that really our beliefs have logical justification after all, in a way it is really an acquiescence in skepticism itself. Speaking in the language of the traditional epistemology, Quine is arguing that we should learn to live with skepticism, by taking pragmatism as a source of solace.

Does pragmatism support naturalism?

Perhaps pragmatism is a source of solace, but is it the right kind? Quine wants to be pragmatist about justification, but at the same time wants to privilege the pragmatic power of science. Every element of his naturalized epistemology has been motivated by the commitment to scientific theory as the highest form of human discourse. Any form of discourse—intensional or modal, etc.—which does not fit into the working vocabulary of rigorous natural science has been disqualified as inadmissible for philosophic purposes.

But if Quine is not attempting to offer a traditional, logical justification of scientific theory, and is merely pursuing a pragmatic justification, what manner of pragmatic justification privileges natural science over the many other modes of human discourse? Richard Rorty (1979, 171) raises a similar question:

> Quine, after arguing that there is no line between science and philosophy, tends to assume that he has thereby shown that science can replace philosophy. But it is not clear why natural science, rather than the arts, or politics, or religion, should take over the area left vacant.

Later he continues:

> The conviction that science differed from softer discourse in having "objective reference" to things "out there" was bolstered in pre-Quinean days by the thought that…there certainly were points of contact with the world in the presentations of sense. This contact…seemed to give science what was lacking in religion and politics—the ability to use contact with the real as the touchstone of truth. The horror which greeted Quine's overthrow of the dogmas…was a result of the fear that there might be no such touchstone. For if we once admitted that Newton was better than Aristotle not because his words better corresponded to reality but simply because Newton made us better able to cope, there would be nothing to distinguish science from religion or politics (269).

Rorty, of course, thinks that there *can* be no principled distinction between the pragmatic value of science and that of arts-politics-religion. He thinks that Quine (and others) have shown us, inadvertently, that some doctrine is acceptable just in case it is consistent with the standards—scientific or otherwise—of our cultural peers. And Rorty is not the only one to have seen an affinity between his views and Quine's. Note the following from Hilary Putnam (2004, 65):

> [I]f neither criterion has any pretension to providing a sense in which our propositions are capable of mapping the behavior of specific hunks of reality…then valorizing prediction of nerve-stimulations over "coping" broadly construed is (as Rorty tirelessly points out) utterly arbitrary. Quine, it seems to me, gave up realism without noticing that he did, because he thought that as long as he valorized scientific discourse above all other discourse, this *made* him a realist…. "Naturalism" is unstable indeed if it slides so easily into Rortian antirealism.

Quine, of course, would likely disavow the extreme cultural relativism espoused by Rorty. The question, of course, is how he would then propose to privilege the pragmatic authority of *Naturwissenschaften* over *Geistewissenschafen?* This is, of course, a question that would need to be addressed within the confines of science itself—or at least in whatever immanent theory we happen to find ourselves in. Presumably Quine could present a formidable argument showing that only the scientific pragmatic virtues of simplicity, conservatism, and empiricism are conducive to predicting further sequences of sensory stimulation. But what if other pragmatists are not so interested in predicting their sensory sequences? Surely prediction enables control, but what if other pragmatists are interested in coping with life through other means than controlling life's experiences? If Quine wanted to give a definitive answer to this question, he would need to find a way of naturalizing the norms of prediction and control. Given the material I have already discussed in chapter 3 concerning the difficulty of naturalizing normativity, Quine would face an uphill battle.

Of course this criticism might be besides the point. Probably all Quine ever intended by his naturalism was to exalt science as the ultimate source of *truth*, not the ultimate source of coping. Whether and to what extent truth is to be understood as a norm is, perhaps, a secondary question. The primary question is: whatever our reasons for wanting the truth, how do we find it? If Quine could then argue that whatever form of discourse permitted us to predict our sequences of sensory stimulations *also* thereby permitted us to make conclusions about the truth would be the discourse of interest to the philosophical naturalist. We already recognize that wherever questions about unobservables or other theoretical matters are underdetermined by the evidence, we must rely on pragmatic virtues which are not themselves linked to truth. This much pragmatism possibly threatens scientific realism, though not necessarily realism *per se*. But there is still the possibility that at least our observation sentences, the statements which report our sensory stimulations, may be adequately linked with truth. The question is how.

In a separate article, Putnam (1995) presents some serious difficulties for the Quinean attempt to formulate a truth-linked pragmatism, even about observation sentences. In the remainder of this section, I will present some highlights of Putnam's concerns. I will not have occasion to consider objections to Putnam, and for this reason I do not mean to endorse him wholeheartedly.

Putnam begins by considering Quine's reliance on a deflationary notion of truth, as explicated by a Tarski-style true-in-L predicate, according to which "Snow is white" is true-in-L if and only if *snow is white*. This is a notion of truth that we are supposed to be able to rely on in spite of problems of referential inscrutability. Of course the Tarski definition applies directly only when the object language is contained in the metalanguage. If we want to state the true-in-German definition for "Schnee ist weiss," then we need to translate "Schnee ist weiss" from German with its English equivalent, and say "Schnee ist weiss" is true-in-German if and only if snow is white. But then we need to know that our translation is correct, which Quine of course says there is no fact of the matter to determine. Therefore sentences in German are true or false only in relation to a translation manual with English. This quandary even applies to our home language if we view it as just a set of speech dispositions. But as we have already explored, we can take solace in acquiescing in our mother tongue, taking our words as face value. We can assert wholeheartedly that "Snow is white" is true if and only if *snow is white*. So only in acquiescing in our own language do we grasp that our words refer: but where there is no theory, there is no determinate reference. Apart the context of our language, there are no determinate reference relations between our words and the world.

Much of this we have already explored. But Putnam considers the natural question one often encountered after examining the Tarski truth definition: how are we ever to *know* whether or not snow is white? The usual Quinean response is that this is a question for epistemology, not for a theory of truth. Whatever leads us to say *snow is white* is what leads us to accept that "snow is white" is true. But Putnam asks, "how can there be an epistemology in connection with truth if there is no truth to be found?" (339). He notes a shortcoming in the analogy Quine draws between the relativity of truth to a translation manual and the relativity of position to a coordinate system. We can give an absolute statement of position relative to a coordinate system even if we are not using that coordinate system, but we cannot even give an absolute statement of truth relative to a translation manual if that manual is not one in which we have acquiesced. (He gives an example using a Martian linguist whose meta-language is not necessarily ours.) Noting this, he says:

> The only solution consonant with Quine's general position that I can see would be to abandon the geometrical analogy and to say that in the case of my own language, calling a sentence 'true' is doing no more than reaffirming the sentence. I am not ascribing any property, not even a relative property, to the sentence when I say that it is true. I am just assenting to the sentence. Quine himself puts the matter this way when he says that 'to say that a sentence is true is just to reaffirm the sentence.' On this interpretation, to say, as Quine does, that there is only 'immanent truth' is—as close as makes no difference—to say *il n'y a pas de hors text* [there's nothing outside the text]. (341)

If Putnam is correct, and not even such a thing as an absolute relative truth predicate, it does sound like truth is no predicate at all, just a device used for semantic ascent. Perhaps a device like this is all we need if we are interested in describing our agreement or disagreement with others, and other pragmatic functions often cited by advocates of the deflationary conception of truth. But it would seem that if we want to be able to show why science rather than art-politics-religion is the preferred mode of human discourse, we will need a truth *predicate* that does more work than that. We want to be able to say, quite abstractly, that there are certain modes of discourse which produce *true statements* about the world, in virtue of features of that discourse which are *truth conducive*. Part of the reason we need to be able to state it abstractly is precisely because we do not know which examples are supposed to fall under the extension of "truth" to begin with. If we are approaching a new question to answer, and we have two competing worldviews telling us how to answer it, as epistemologists we want to know which worldview has the *general features* that make it conducive to truth. We want to be able to say that one, rather than the other, derives from the senses, and that the senses are in causal contact with the world. It is not enough just to be able to say that one worldview produces various sentences, and then affirm those sentences. Often competing worldviews agree on a lot. Even Quine's naturalized epistemology attempts to do more than this. He notes that science makes an important connection between theory and the senses. But why is any special connection needed at all? Why do we even need any method for pursuing truth, if truth is not any distinctive property?

I offer Putnam's concerns above simply as evidence of a difficulty with Quinean pragmatism that *must* be settled if naturalized epistemology is to retain its exclusive reliance on natural science. Because I am not presenting a comprehensive case for Putnam's view, I take it that in this chapter I have *not* proven that Quinean naturalism is inadequate on its own terms. (It seems to me possible that a deflationist could address some of the questions I raise above about how deflationism could characterize the general property of truth-conduciveness needed by epistemology.) I will have shown only two things up until this point: 1) Traditional epistemologists who are interested in justification of the truth of scientific claims beyond those immediately related to sensory stimulations will see Quinean naturalism as a form of skepticism, and 2) Quinean naturalism faces serious difficulties even in formulating an epistemology that links sensory stimulations to truth. Both of these points taken together will give us reason, if not to reject Quinean naturalism outright, then at least to decide if it is well-motivated. If we do not like the idea of giving up traditional epistemology, and if we are not up to the difficult task of solving Putnam's problem, we should decide if we really have to. In the final section of this chapter, I will explore the source of the principles which motivate Quinean naturalism. This will enable us, in our final chapter, to assess the source of these principles, and decide if we really need to abandon traditional epistemology and be Quinean naturalists after all.

Proximate sources of inscrutability and indeterminacy

In this final section, I would like to show how two of the most important sources of Quine's rejection of traditional epistemology—his inscrutability of reference and indeterminacy of translation theses—ultimately reduce to another: his underdetermination thesis, or more broadly the confirmation holism from which underdetermination immediately springs. This will not be an exercise in reduction for the sake of reduction, because I think that in the end, showing the ultimate sources of Quine's rejection of traditional epistemology (and acquiescence in skepticism, traditionally understood) will provide us with a clue for how to circumvent that position. Furthermore, showing the roots of Quine's theses in underdetermination will help solidify my claim that there is a sense in which the proposal to naturalize epistemology represents an acquiescence in skepticism, traditionally understood.

Let us begin with a brief review of how Quine wields his various negative theses (underdetermination, inscrutability, indeterminacy) as weapons against traditional epistemology. This is important to do if we are to show that Quine's case for naturalism is not just a "pessimistic induction" about the failures of past epistemologies. That is certainly a part of his case, but not the whole. In "Epistemology Naturalized," Quine (1969a, 69–70) divides epistemology into two projects, the "doctrinal" and the "conceptual". The first of these is concerned with justifying our knowledge (for empiricism, in sensory terms), whereas the second is concerned with clarifying the meaning of key concepts (which relates to the doctrinal project insofar as translation of obscure truths into clearer ones can help justify them). Quine notes that the doctrinal project had failed in empiricism at least by the time that Hume had formulated his problem of induction (this is what Quine means by the "Humean predicament") (71–2). The conceptual project, however, was kept alive as long as various empiricists attempted to reduce talk of "bodies" to sensory terms, although again, Quine thinks the most advanced attempt (Carnap's) was a failure. But this is not the whole story.

Critics such as Kim (1988) have argued that in citing these failures as reasons to abandon traditional epistemology *in toto*, Quine has unjustifiably discounted the possibility of new developments of either the doctrinal or conceptual projects that could fare better than earlier attempts. In chapter 3, however, I have argued that *each* of the new developments proposed by Kim could fall prey to one or another of Quine's negative theses. Kim first focuses on the doctrinal project, listing probabilism, coherentism, and externalism as projects worthy of exploration, under the auspices of traditional epistemology. The first of these is ruled out by underdetermination, which implies that not even the belief in physical objects is more probable than a belief in the gods of Homer (the sole difference is pragmatic). Traditional coherentism is no option for Quine, because of his rejection of analytic *a prioricity*, which is needed by coherentism to assign initial plausibility to certain beliefs to prevent the justification of coherent fantasies. After having examined Stroud's critiques of Quine, which in effect attributed to Quine an externalist account of justification, we can now see in particular why externalism offers no hope for genuine justification on Quine's view. In addition to this, any of these views which attempt to *analyze* "justification" in terms of probability, coherence, or reliability of belief-formation, are to be repudiated out of hand by Quine, insofar as he rejects "conceptual analysis" as a proper philosophic methodology, because of his rejection of meaning and embrace of indeterminacy. Even Kim's own preferred concept of philosophic methodology, supervenience, is likely to be dismissed on naturalistic grounds, owing to its reliance on suspicious modal concepts.

As I have argued in earlier sections, Quine's negative theses, in particular the inscrutability of reference and the underdetermination of theory by evidence imply skepticisms of their own. It is thus not surprising that so wielded, they should also rule out alternative doctrinal projects as proposed by Kim and others. So what is the source of these negative theses? I shall begin with the indeterminacy of translation, and show how it reduces either directly to an underdetermination thesis or to the inscrutability of reference, which then itself reduces to another underdetermination-style thesis. The ways in which indeterminacy and inscrutability reduce to underdetermination are not precisely reductions to the underdetermination thesis, per se (the one about theory and evidence), because strictly speaking, translation manuals and reference schemes are not theories. Thus Quine and defenders of Quine have long held that there are important differences between indeterminacy and inscrutability on one hand, and *the* underdetermination thesis on the other (Quine 1969, 303; Quine 1970, 80; Quine 1981b, 23; Quine 1987, 10; Quine 1992, 101; Gibson 1986; Peijnenburg and Hünneman 2001). That is true, there is a distinction. But the distinction itself is a product of the arguments for indeterminacy and inscrutability, which themselves turn on a premise reminiscent of the underdetermination thesis.

First, let's examine the indeterminacy of translation—a device that could be exploited by Quine to reinforce his critique of *anyone's* attempt to define conditions for knowledge or justification, analytically or otherwise. In his "On the Reasons for the Indeterminacy of Translation" (1970), Quine outlines two separate arguments for indeterminacy. The first, which he denotes by "pressing from above," is from the underdetermination of physical theory. The second, "pressing from below," is from the inscrutability of reference. Like Quine, I will focus on "pressing from above" first.

In his first argument, Quine takes it for granted—and expects that others will concede—that physical theory is underdetermined, not only by past observations, but by all possible observations. Turning to the question of translation, Quine argues that it is easy and objective enough to match our observation sentences with those of the foreigner, but that translating theoretical sentences requires projecting analytical hypotheses. It is with analytical hypotheses for translating *theory* that the bulk of indeterminacy enters, because:

> [N]ow the same old empirical slack, the old indeterminacy between physical theories, recurs at second intension. Insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences. If our physical theory can vary though all possible observations be fixed, then our translation of his physical theory can vary though our translations of all possible observation reports on his part be fixed. Our translation of his observation sentences no more fixes our translation of his physical theory than our own possible observations fix our own physical theory (1970, 180).

This confirms what one might have suspected from reading *Word and Object*, that the indeterminacy of translation follows from the fact that translation manuals are underdetermined by the relevant naturalistic facts, sensory stimuli and speech dispositions:

> Yet one has only to reflect on the nature of possible data and methods to appreciate the indeterminacy. Sentences translatable outright, translatable by independent evidence of stimulatory occasions, are sparse and must woefully under-determine the analytical hypotheses on which the translation of all further sentences depends…. There can be no doubt that rival systems of analytical hypotheses can fit the totality of speech behavior to perfection, and can fit the totality of dispositions to speech

> behavior as well, and still specify mutually incompatible translations of countless sentences insusceptible of independent control. (1960, 72)

But Quine is quick to reassure us that this does not mean that indeterminacy of translation is an *instance* of the underdetermination of our theory. A translation of another's sentences is not just a theory we have about the meaning of their sentences. This is in part because Quine refuses to accept that there are such *things* as meanings of individual sentences, for the familiar reasons stemming from combining confirmation holism and verificationism (1969a, 80–81).[45] Translation is not a new theory using a new set of theoretical terms. It is simply the matching of one's own home theory with that of the foreigner, or as George Romanos puts it, a "way of reading this theory…into the language he is investigating" (1983, 181). While there is a fact of the matter according to which underdetermined theories are true or false, there is no fact of the matter according to which rival but incompatible manuals of translation are correct or incorrect (Gibson 1986, 151–2).[46] When we engage in translation, we presuppose our theory of physics—call it A—and hold it as true, even though a variety of incompatible theories (A, B, C, etc.) are empirically equivalent to it and equally underdetermined by the evidence. Presumably we have used pragmatic factors (such as simplicity) to choose our theory, but when it comes to reading this theory into the speech disposition of others, the same pragmatic factors do not dictate what we attribute to them. Simplicity may dictate assigning a *false* (by our lights) theory, B, to the foreigner, for example. There are other possibilities (Quine 1970, 180). But

---

[45] To the extent that the underdetermination thesis *itself* relies on confirmation holism, the indeterminacy of translation thesis is really a *triple* iteration of confirmation holism: one to say theory is underdetermined, another to say translation manuals are underdetermined, and a third to say that because there are no things that are the meanings of individual sentences, there is therefore no fact of the matter for translation manuals to concern themselves with. This makes one think confirmation holism is rather important to Quine's philosophy!

[46] Peijnenburg and Hünneman (2001, 20-4) have argued that Romanos' and Gibson's explanations for the difference between underdetermination and indeterminacy are themselves different. I don't see any grounds for this. On this view, Romanos argues for indeterminacy simply from the fact that there is a "double underdetermination" involved in translation. On its face, though, this explanation gives no reason for thinking that indeterminacy should follow from underdetermination. I think it only follows given the assumption that the second iteration of underdetermination has no fact of the matter to appeal to, i.e., Gibson's understanding. But I think Romanos is assuming this explanation implicitly, when he says that a translator is simply taking his own theory and trying to put it into new notation (184). This implies, I think, that there are no new facts beyond the "notation," and of course in radical translation the "notation" we are dealing with is the set of speech dispositions, etc.

these factors are *only* pragmatic. Because there is no reason to believing in the meanings of individual sentences, there is no reason to believe we are dealing with any new facts of the matter beyond observations and speech dispositions.

But even though indeterminacy is not an instance of the ordinary underdetermination of physical theory by evidence, this does not mean it is not an iteration of the same old underdetermination argument. It is particularly revealing, I think, that Quine suggests above that the candidates for translating our theory into the language of others are simply expressions of the rival theories we ourselves consider as being underdetermined by our evidence. In other words, if physical theory itself is not underdetermined, there would be no way of conceiving of alternate translation manuals, and translation would not be indeterminate. Thus Quine states, "What degree of indeterminacy of translation you must then recognize…will depend on the amount of empirical slack that you are willing to acknowledge in physics" (1970, 181). Quine then considers various degrees to which one might accept underdetermination in physics, with corresponding degrees of indeterminacy in translation. But by this logic, if we were not to recognize any underdetermination in physics, there would likewise be no indeterminacy of translation. This is a point that seems to hold even if we were to fail to specify a "fact of the matter" that translation would concern. Perhaps there would be no such fact, and translation would only be a matter of reading one's only possible theory into another's language. The best one could argue at this point would be a non-cognitivism about translation, not an "indeterminacy," on the grounds of Quine's argument combining verification and confirmation holism (if meaning is the method of confirmation, and only whole theories, not sentences are confirmed, then only whole theories, not individual sentences, have meanings).

So I think it is clear how Quine's "pressing from above" argument for indeterminacy relies on the underdetermination thesis. What of "pressing from below," the argument from the inscrutability of reference? This, of course, is the argument regarding the reference of individual terms, like "gavagai." Does "gavagai" refer to rabbits, or rabbit stages, or further bizarre slices of reality made possible by proxy functions like "cosmic complements," etc.? Quine thinks that inscrutability could also help establish the indeterminacy of meaning/translation, but he doesn't seem to think it is as definite as "pressing from above." Indeed Quine thinks observation sentences may be translated objectively (or at least close to objectively), even though their terms, when divided, are inscrutable. He says that the point of the "gavagai" example "was aimed not at proof [of the indeterminacy thesis] but at helping the reader to reconcile the indeterminacy of translation imaginatively with the concrete reality of radical translation" (1970, 182), though it is not entirely clear what he means by this. He even suggests that an allegedly undebatable example of inscrutability of certain features of Japanese grammar does not imply any indeterminacy of translation (1970, 182).

Whatever Quine says about the link between inscrutability and indeterminacy, it is difficult to keep the issues of inscrutability and indeterminacy separate. To begin with, if what looks like an individual term is *itself* to be taken as a one-word sentence, then indeterminacy in regards to that single term would map onto indeterminacy in translating it as an entire sentence. Also, the question of whether to regard it as a term or a sentence itself renders the sentence indeterminate. Granted, the cases in which inscrutability does not seem to imply indeterminacy are those cases (the translation of observation sentences) which Quine originally believes to be objectively translatable. But in later days, Quine's commitment to the objective translatability of these sentences becomes less forthright. He begins to suggest that translators must rely on subjective-sounding "empathy" in order to ascribe observation sentences to others, no longer being able to rely on an objective knowledge of their stimulus meanings (Quine 1992; 1996). Observationality becomes a more graded notion, and for this reason the extent of indeterminacy spreads. For all these reasons, it seems at least worth exploring inscrutability as a source of indeterminacy, which is probably why Quine includes it in "On the Reasons for Indeterminacy of Translation," and says "pressing from below" consists of "whatever arguments for indeterminacy of translation can be based on the inscrutability of terms" (1970, 183).[47] Finally, even if inscrutability does not imply indeterminacy, it is worth discussing here, because as I argued above, inscrutability is still an independent source of skepticism, and we are looking for the roots of Quine's rejection of traditional solutions to skepticism.

One recent commentator, Nimtz (2005, 4), summarizes the argument for inscrutability of reference as follows:

> (a) Semantic properties are exclusively determined by A facts.
> (b) For any language L, the totality of A facts is compatible with indefinitely many radically different interpretations.
> (IR) [Therefore,] [f]or any language L, there are indefinitely many radically different yet equally correct interpretations.

One may, of course, object to Quine's list of A facts as stemming from an impoverished behaviorism. But I think even more crucial to Quine's argument here is not his assumption about what counts as naturalistically acceptable, but his move from "*is compatible with indefinitely many interpretations…"* to "*there are indefinitely many …correct interpretations.*" The slide seems almost imperceptible, but makes sense if we assume that an interpretation's compatibility with the A-facts is sufficient to make the interpretation *correct*.

Leaping from mere compatibility with the evidence to equal correctness should remind us of the underdetermination thesis. The underdetermination thesis is also about equal "correctness," equal correctness with respect to evidence. According to this thesis, the evidence does not decide between empirically equivalent theories, because the theories are each compatible with the same evidence. As we shall see in chapter 6, this usually stems from the assumption that to for a theory to have the same logical consequences as another just is to have the same empirical support, to be equally "correct." Of course the inscrutability thesis is not a thesis about theories and their evidence, but about the reference of our terms. Nevertheless, the *argument* for this thesis implicitly relies upon a view about evidence,

[47] For further reasons to see a relation between inscrutability and indeterminacy, see Kirk (2000, 165), and Orenstein (1997). See also Quine (1960, 71-2).

because deciding that an interpretation about reference is "correct" is just the same as deciding if we have evidence for thinking that some reference scheme holds.

Consider, for example, how asking "is this the same rabbit as that?" would settle whether "rabbit" refers to "rabbit" or "rabbit stage." If the subject were able to give a determinate answer "yes" while pointing at a rabbity patch, we would know that his reference scheme extended beyond the immediate moment, referring to an enduring object, rather than just a rabbit stage. But Quine believes that there is no way for the subject to answer with a determinate "yes" or "no" to our question, because the question itself could translate as "is this the same rabbit?" or as "is this the same succession of rabbit stages?" I think the same style of argument is at work in arguments for inscrutability that stem from proxy functions: using these cases, we conclude that we may indirectly ostend many different types of objects, but the suggestion that there is no fact of the matter concerning which we indirectly ostend is still made by reference to the limited facts about what we directly ostend. As Quine puts it, "even the pointing out of an inscription is no final *evidence* that our talk is of expressions and not of Gödel numbers" (1969c, 44, emphasis mine).

So the argument for the inscrutability of reference thesis says that if we have a theory about the reference of another's terms, the evidence we might adduce for such a theory underdetermines any of the possible reference scheme attributions. So, Quine is not saying that to *have* a reference scheme is something that requires evidence, or that one's *possession* of a reference scheme is therefore underdetermined by evidence. The conclusion of his argument is just that there are no facts or evidence that determine such a scheme. But *his argument* for this claim is one that *does* rely on the use of evidence, even though in the end he would say that even the attribution of a reference scheme not a theory of our own.[48] What Quine is saying when he argues that "rabbit," "rabbit stage," etc. are

---

[48] It is not a theory of our own for probably the same reason that a translation manual is not thought to be a theory. "Reference" is not a theoretical term that refers to entities called "references," or at least he would say we have no evidence that such a term could be used to explain, because a fixed amount of evidence is consistent with a variety of divergent reference schemes.

compatible with facts about direct ostension of *gavagai*, is that *if* we take any of these as the reference scheme, *then* we will ostend identical stimuli and answer identical questions about reference. If we are referring to rabbits, then we will ostend the exact same stimuli and answer the exact same questions, as we would if we are referring to rabbit stages. Likewise when he argues that theory is underdetermined by evidence, his argument is that *if* theory A is true, *then* such and such empirical consequences will obtain, and *if* theory B is true, *then* the same empirical consequences will also obtain, so both theories are equally well-supported by the evidence and therefore underdetermined by it.

Indeterminacy of translation, inscrutability of reference, and underdetermination of theory by evidence each have apparently skeptical consequences when they are raised against traditional epistemology, as I have argued both here and in chapter 2. If my argument in this section is correct, and both indeterminacy and inscrutability reduce to the underdetermination-style theses, this suggests that it is the ultimate source of the skepticism. I don't think it should come as much of a surprise that underdetermination has apparently skeptical consequences, because some of the most prominent skeptical arguments in the history of philosophy have been underdetermination arguments. As Okasha (2003) observes, even Descartes' evil demon argument works by noting that our experience is compatible with a variety of empirically equivalent alternatives: that there really is a fire in front of us, that we are dreaming that there is a fire in front of us, that an evil demon has made us believe that there is a fire in front of us, etc. So, seeing Quine's negative theses as rooted in the underdetermination thesis helps us to see just how much affinity there is between naturalism and skepticism. Quine's pragmatism may allow him to explicitly disavow skepticism, by permitting him to rely on extra-empirical criteria to decide between empirically-underdetermined rivals, but to the traditional epistemologist, this just is a concession to skepticism.

Conclusion: Reciprocal containment revisited

Quine thinks epistemology is prompted by skeptical doubts. This essay was also prompted by a doubt, one concerning how Quine intends to respond to skepticism. Critics of Quine take him to be trying to refute it, but failing; defenders seem to agree he is trying to refute it, and believe he succeeds. I have argued that neither party completely appreciates Quine's total project. Quine is not trying to refute the skepticism that concerns the traditional epistemologist. He takes the skeptical problems generated by traditional epistemology as basically irresolvable on their own terms.

Naturalized epistemology is not a refutation of skepticism, but an accommodation to it. It is the attempt to explain the origin of our beliefs, even the origin of what we take to be true beliefs, under the assumption that these beliefs did not arise from anything like awareness or logical justification. This is the full meaning of Quine's idea that doubt is what prompts epistemology. The language of "prompting" is quite appropriate here, because we epistemologists, for Quine, are just like our behavioristic subjects. We too are prompted by stimuli in our environments. In this case, the stimulus is our own skeptical doubts. Doubt prompts naturalized epistemology, not because we wish to erase this doubt and find secure foundations for our knowledge (which are impossible), but because we take this doubt for granted and want to understand how we were ever able to erect an edifice of theory *in spite* of having no secure foundations. From Quine's perspective, it is an impressive feat our species has achieved—the erection of modern science on the basis of sloppy analogies, wistful symmetries, and various other doctrines "conceived in sin."

There is one last doubt prompting our further inquiry: if underdetermination is what prompts the project of naturalized epistemology, what are we to make of the fact that it is only under the assumption of naturalism that underdetermination can be taken seriously in the first place? Does this make Quine's overall project circular? This might make things easier for the critic of Quine, if the circle is vicious. But as Quine has repeatedly urged, an epistemology that rejects foundationalism does not need to treat circularity as vicious. How, then, can critics break into the coherent circle of Quine's theory and undermine it, if there is no starting point to attack?

First we need to review the sense in which the underdetermination thesis itself has its origins in naturalism. We have already seen the dominant way: Quine's naturalist looks first to an empiricist psychology, which points to the exclusive informational relevance of stimuli on our sensory surfaces. The naturalist concludes that more than one scientific theory or reference scheme is compatible with the sparse stimuli of the sensory surfaces. This is not the entire story, of course. A scientific story must be told about how sensory stimuli coalesce with social reinforcement in order to give rise to observation sentences. Another scientific story must be told about how observation sentences relate to theory, via Quine's so-called "observation categoricals." This last is Quine's particular version of the hypothetico-deductivist view of confirmation, a view which he takes to be grounded not primarily in the content of science, but in the facts of scientific practice. And as we shall see in the next chapter, it must be taken for granted that hypothetico-deductive connections are the exclusive links between theory and observational evidence.

Each of these naturalistic sources of the underdetermination thesis represents an aspect of the containment of epistemology in natural science. So there is a sense in which the underdetermination thesis already presupposes naturalism. Quine is quite content with this, however, because of his doctrine of "reciprocal containment": he also takes natural science itself to be contained in epistemology. It is our empiricist epistemology that tells us to look to science as our highest source of knowledge in the first place (and it tells us that, of course, because it in turn is contained in a natural science that tells us that our only contact with the world is through our sensory surfaces). So it is Quine's doctrine of reciprocal containment that best represents his anti-foundationalism. Neither science nor epistemology is a "starting point" for the other. Each is viewed as an instance of the other depending upon perspective.

But the fact of reciprocal containment does not imply that naturalized epistemology is enclosed by an impenetrable circle. We can still ask questions about the specific content of the science assumed by the naturalized epistemologist and about the specific commitments of empiricist epistemology. We might find that even if we, as naturalists, take all of our science seriously, none of it implies the underdetermination thesis or requires that we become naturalized *epistemologists.* And we might find that even if our only contact with the world is through our senses, this is fully consistent with a traditional approach to epistemology.

There are several ways to challenge the content of Quine's science, in ways that show that real science does not contain his particular version of naturalized epistemology. We will do much of this in the next chapter, primarily by challenging the hypothetico-deductivist account of confirmation by reference to historical scientific practice. We will also challenge the traditional empiricist accounts of sensation and concept-formation which impoverish theories of confirmation, eliminating the possibility of inductivist foundationalism. Apart from this, there are many other ways in which one could challenge the psychology of language-learning behind Quine's account, by challenging his "holophrastic" account of observation sentences (Bloom 1973), and his social-linguistic behaviorism (Nelson 1988; Bloom 1993; Modee 2000). Going further, one could even challenge Quine's central dogma that extensional language is the exclusive language of science (Hookway 1988).

As we shall see, this angle of attack on Quine's system resembles the strategy of defusing the skeptic's *reductio,* which Stroud and others mistakenly attribute to Quine himself. If arguments for skepticism are of the *reductio ad absurdum* form, then by showing that a wider body of scientific evidence does not contain absurdities like the underdetermination thesis, we acknowledge the importance of science without showing that it implies the need to abandon traditional epistemology. In short, it may well be that science tells us that our only source of information is through the senses, but this does not imply that science obligates us to be *naturalized* empiricist epistemologists. What it *means* for the mind to be in contact with the world by the senses is a matter of some controversy, one that probably requires some philosophic interpretation. Contact with the world through the senses may not be as impoverished and conducive to skepticism as naturalized epistemologists might think.

So even if there is a sense in which natural science contains epistemology, the points above suggest it may not contain *Quine's* epistemology. We can also qualify the converse containment, of science in epistemology, in a way that calls the total package of naturalized epistemology into question. It may well be that we celebrate the empirical method of science by reference to prior empiricist epistemology (which is, in turn, endorsed by science). But there are also significant differences of opinion over *how* science should be celebrated. It may well be that, if empiricism tells us that it is only through the senses that we acquire information about the world, then we should recognize science as the *highest* form of knowledge. It is, after all, the most systematic evidence-driven discipline. But recognizing science as the highest form of knowledge does not imply recognizing it as the only form, or even as the only *empirical* form. There may also be commonsense empirical knowledge on which science is built. In fact I will argue in the next chapter that to the extent that science undermines the crude empiricism of the naturalized epistemologists, it also makes room for the epistemological relevance of first-person pre-scientific experience. Once again, that leaves room for more traditional approaches to epistemology, which, if not *a priorist* are at least not naturalized.

The last point, about the possibility of pre-scientific knowledge, would call into question that central doctrine of naturalized epistemology, that science was "conceived in sin," and is therefore in need of some non-logical, psychological explanation. To show that this "original sin" approach to knowledge may in fact be a myth, however, it is our obligation to examine the science that demonstrates this. As I have argued, there are several ways to show that science does not contain the kind of epistemology Quine champions. In the next chapter, I will focus on just a few of these issues. To assess the underdetermination thesis on which both inscrutability and indeterminacy rest, I will challenge both the hypothetico-deductivist and anti-inductivist views that many empiricists before

Quine took for granted. If these challenges hold water, we will see that it is Quine’s acquiescence in skepticism which is truly the “overreaction.”

CHAPTER 6

ESCAPE FROM THE HUMEAN PREDICAMENT

As we discovered in the previous chapter, a number of critics contend that Quine has no serious reply to the skeptic, and that his view itself leads to skepticism. Quine has an interesting response to this attack: because skeptical doubts themselves arise within science, we can conclude that the skeptic is "overreacting." The aim of this chapter is to propose a new anti-skeptical strategy inspired by this response, one that can even be used against skeptical elements of Quine's own philosophy, and at least in part against other prominent forms of skepticism in the empiricist tradition.

One of Quine's examples of the scientific nature of skeptical doubt concerns forms of the argument from illusion. Quine points out that these cases presuppose an understanding of "veridical" perception of reality, in contrast to which some cases are understood as "illusory." Another example is the doubt stemming from the discovery of the retinal image, and the worry about how three-dimensional awareness can result from projection on a two-dimensional surface. Quine points out that even this worry presupposes an understanding of a three-dimensional eye in a three-dimensional world.

As I have noted in chapter 5, it looks like Quine does not take these points to establish that skepticism is somehow incoherent. He writes that the skeptic "is quite within his rights in assuming science in order to refute science; this, if carried out, would be a straightforward argument by *reductio ad absurdum*" (1975b, 68). Quine's claim that he is only criticizing the skeptic's overreaction (rather than any incoherence) leads critics such as Stroud (1984) and Williams (1996) to suggest that Quine is trying to use science to defuse the skeptic's *reductio*: if the skeptic has the right to assume science at the outset of a *reductio*, then anti-skeptic also has the right to show how *full* consideration of science blocks this *reductio*, by showing that skeptical absurdities are merely apparent, arising only from scientific assumptions made in isolation from others.

But in chapter 5, I suggested that Stroud and William's interpretation of Quine is incorrect. If Quine intended to respond by defusing the skeptic's *reductio*, then by his own terms it would be an inadequate response, because he takes some of the most important discoveries of the naturalized epistemologist to be that reference is inscrutable and theory underdetermined by evidence. These theses by themselves seem to engender skepticism. If anything this makes the findings of science sound more, not less, absurd. As it happens, I argue that Quine does not mean to dissolve skeptical doubts so much as suggest that they do not matter, because we can always "acquiesce" in our own language and theory, and find consolation in pragmatic rather than logical forms of justification.

However Quine's pragmatism is not necessarily satisfactory. A traditional epistemologist will of course lament the abandonment of traditional goals of explaining the logical justification of science, and it is also not clear if a pragmatic criterion can succeed on its own terms in privileging science in favor of other forms of useful human discourse. So it is important to explore whether the skeptical theses that motivate Quine's pragmatism are true in the first place. At the end of the last chapter, I argued that the indeterminacy of translation and inscrutability of reference theses could each be reduced in large part to versions of the underdetermination thesis. Now indeterminacy, inscrutability and underdetermination are not the only points in the case for naturalizing epistemology. As we saw when we examined Kim's proposed alternatives to naturalized epistemology, behaviorism and extensionalism also help to eliminate traditional epistemological projects and help leave naturalism as the only viable proposal. In the previous chapter, I mentioned in passing strategies that might help undercut both behaviorism and extensionalism, but said I would focus on the underdetermination thesis in the current chapter. I will focus on it by asking the following question: Even if Quine did not intend to dissolve skeptical doubts by pointing to the scientific nature of these doubts, what if *we* were to use that strategy against the very skeptical theses, in particular against the underdetermination thesis, the embrace of which necessitates Quine's acquiescence in skepticism? What if we can abandon the underdetermination thesis itself by determining that, while the thesis appears to arise from some important scientific facts, it dissolves in the light of the wider context of scientific findings?

In this chapter, I hope to show how this strategy can be deployed against the underdetermination thesis by drawing on existing anti-skeptical arguments from a number of philosophers. Drawing on material from Larry Laudan and Jarrett Leplin among others, I present the assumptions about the nature of scientific practice behind the underdetermination thesis, and show what further scientific assumptions may be appealed to in order to defuse it. In my second section, drawing on material from John Norton, I examine a strategy for dissolving another source of radical skepticism, the classical Humean problem of induction, based on consulting the wider context of scientific practice. In my final section, drawing on material from J.J. Gibson and Jesse Prinz, I show how a final anti-skeptical strategy can be deployed against the anti-foundationalism which motivates inductivist skepticism. In each case, I show how skeptical doubts can be opposed by the free use of science within the scope of skeptical reductios, allowing us to escape from the "Humean predicament."

There will be just one catch. Finding the scientific source of doubts is sometimes only a partial solution to skeptical arguments. I say that because in some cases, I think skeptical doubts arise from scientific evidence only in conjunction with particular philosophical assumptions, and these assumptions are not always derived directly from science themselves. As we move on to more and more fundamentally skeptical problems, such as the problem of induction and anti-foundationalism, the doubts that arise will become more philosophical in nature. In these cases, I will try to identify and, in some cases, refute erroneous philosophical assumptions. But I do not need to refute them completely to make my point: the presence of purely philosophic, non-scientific assumptions in the case for naturalizing epistemology shows, I think, that Quine is not entirely correct that skeptical doubts are always scientific ones. But if that is the case, it follows that to the extent that he has relied on any of these doubts to motivate his naturalism and pragmatism, he has also relied on non-naturalistic presuppositions, which would undermine the idea that philosophy can ever be entirely continuous with natural science.

This discussion will, therefore, address the "reciprocal containment" of science and epistemology in three ways. First, showing that the underdetermination thesis and its assumptions dissolve in light of a wider context of scientific evidence will help to show that naturalized epistemology is not contained in natural science in the way Quine has suggested. Second, showing that the case for naturalizing epistemology relies on non-scientific philosophical assumptions will show that naturalized epistemology is not contained *exclusively* by natural science. This means that we can accept science without feeling obliged to abandon epistemological tradition. Finally, as promised in the last chapter, our examination will suggest that while an understanding of the value of science is contained in an empiricist epistemology, a reformed empiricism (one that conforms to wider context of scientific information to be examined) need not champion scientific empirical knowledge as the *only* kind of empirical knowledge. There will be room for first-person pre-scientific knowledge, of the sort that will underpin a foundationalism that challenges the need for naturalizing epistemology.

Understanding the scientific roots of the underdetermination thesis

What exactly is the underdetermination thesis, and why does Quine accept it? Philosophers sometimes distinguish between a "weak" and a "strong" underdetermination thesis (see Newton-Smith 2001). The weak version asserts that it is possible to construct empirically equivalent theories, i.e., theories that have the same empirical consequences or which can be consistent with the same body of empirical evidence. This is usually seen as uncontroversial in comparison with the strong version, which says not only that it is possible to construct empirically equivalent theories, but *also* that these empirically equivalent rivals are also equally well-confirmed (or well-supported or well-justified) by observational evidence.[49] Strictly speaking, I don't think it is proper to call the first thesis an "underdetermination" thesis at all, because it says nothing about *determination* by evidence. The second and distinctive part of the strong thesis does say that. The strong thesis says, in effect: *because* evidence can consistent with multiple theories (there are empirically equivalent rivals), therefore theories are not determined by observational evidence alone (the empirically equivalent rivals are equally well-confirmed by the evidence).

This fuller statement of this thesis takes the form of an argument ("because . . . therefore"), but it is perfectly natural to differentiate certain important conclusions by reference to the distinctive premises that lead to them, particularly when there are different ways to come to the same conclusion. There may be many other philosophers' reasons for thinking that theories cannot be determined by evidence. Traditional Pyhronnian skepticism is one example. The underdetermination thesis comes to same conclusion by way of the point about the existence of empirically equivalent rivals. In attacking the thesis, I will attack the argument implicitly contained within it: the idea that the possibility of empirical equivalence implies the impossibility of evidential determination of theories.

A note on my use of the term "confirmation": In what follows, I am using the term as a generic placeholder term referring to non-pragmatic logical support (whether deductive or inductive). Sometimes advocates of the underdetermination thesis, who will insist that empirically equivalent theories are not logically determined by the evidence, will also say that because of pragmatic differences between the theories, one is still better "confirmed" than the other. For example, in "The Scope and Language of Science" (2004e, 198), Quine says that simplicity, one of his pragmatic virtues

[49] Notice that this problem is much stronger than the traditional problem of induction. It is not a problem about the logical gap between limited observations and the unlimited scope of generalizations. Rather, it is a logical gap between *all possible observations* and theory. According to the underdetermination thesis, no matter how much evidence we collect for a theory, there will always be an empirically equivalent rival that will account for it just as well. Arguing that the problem of underdetermination thesis arises from special problems concerning the scientist's method of hypothesis—which we will shortly examine—Larry Laudan calls Hume's problem the problem of "plebeian induction," whereas underdetermination is the problem of "aristocratic induction" (Laudan 1981, 73).)

of theory choice, in some sense counts as "a kind of evidence;" in "On Simple Theories of a Complex World" (2004c, 211) he suggests that perhaps "the simpler hypothesis stands the better chance of confirmation."[50] So there are pragmatic and non-pragmatic senses of "confirm." In discussing formulations of the underdetermination thesis, I will use the non-pragmatic sense, because that is often how it is described in the literature, including by Quine himself in "Two Dogmas of Empiricism" (1953b). There Quine *denies* that individual statements "admit of confirmation or infirmation at all" and therefore may be accepted as true come what may, or revised at any time, provided that the appropriate auxiliary hypotheses in one's holistic theories are adjusted (41). As we shall see shortly, this provides Quine's most important argument for the underdetermination thesis (the statement of which follows almost immediately in "Two Dogmas").

So even though it is possible to deny that empirically equivalent rivals are necessarily equally well-confirmed in the pragmatic sense, this is consistent with the idea that theory is not determined by observational evidence, which is what makes the underdetermination thesis an effective tool in motivating naturalized epistemology. This is because the idea that the evidence does not help logically (deductively or inductively) determine theory is cause enough for concern among traditional epistemologists. If we hold one theory, while we know another actual or even *possible* rival theory is equally well-confirmed by the evidence, it follows that the first theory is not evidentially superior to the rival, and it would seem to be irrational to accept it (Bergstrom 1993), at least as long as evidence is our only measure of rationality or epistemic justification.[51] Even if we take pragmatic factors such as simplicity or predictive power as further components of rationality, it not uncontroversial to wonder why pragmatic rationality is relevant to the *truth* of our hypotheses—especially because of the difficulty involved in giving an objective account of the pragmatic values involved in this kind of "rationality."

---

[50] See also Bergstrom (1990, 44).

[51] There are further complications for how underdetermination leads to skepticism, depending upon whether or not the original theory and its rival are taken to be logically compatible. For details on how these complications are handled, see Bergstrom (2004) and chapter 5.

Having clarified the meaning of the underdetermination thesis and the meaning of some of its terms, we are now in a position to see why Quine and others think it is true. First we need to consider reasons for accepting the premise that there are empirically equivalent rivals for every theory. Then we need to consider whether the conclusion about evidential determination follows from it.

There are at least two different reasons for believing in the existence of empirically equivalent rivals. These are not always articulated by Quine, but have been made explicit by a Quine commentator, Lars Bergstrom (1990), without Quine's objection (1990). The first reason Bergstrom calls the "argument from logic": "a given set of sentences is implied by different, and even logically incompatible sentences; hence an empirical content is implied by different theory formulations."[52] According to this argument, any premises we can construct which imply the same empirical consequences are empirically equivalent rivals. Because we can construct them for any theory, empirically equivalent rivals exist.

The second reason to believe in the existence of empirically equivalent rivals is what Bergstrom calls the argument from confirmation holism. Confirmation holism is the thesis that Quine, in "Two Dogmas of Empiricism," claims to share with Duhem: the idea that it is only whole blocks of theory, not individual hypotheses, that are subject to confirmation or disconfirmation.[53] Confirmation holism results from the fairly obvious point that hypotheses imply empirical consequences only with the assistance of auxiliary hypotheses, and the point that in the face of new observations, one can always either retain or reject the hypothesis in question, depending upon whether one rejects or retains associated auxiliary hypotheses. Each possible set of hypotheses and auxiliaries consistent with observational evidence is itself a block of theory, and since there are many such sets possible given the observational evidence, there are therefore many empirically equivalent rivals. I note that the argument

[52] One might, for example, argue for "If A then B" with either set of premises: "If A then C, If C then B," or "If A then D, If D then B." Since "If A then B" could be understood as an empirical consequence--Quine himself treats empirical consequences as "observation conditionals" or "observation categoricals"--this shows how the same empirical consequence can derive from different theoretical premises.
[53] It does appear, however, that Quine was originally incorrect to ascribe this view to Duhem. See Ariew (1984).

from confirmation holism is more attuned to scientific practice than the argument from logic, because it suggests that empirical equivalents are generated not by completely artificial means, but from transformations of existing bodies of scientific belief. Quine himself (1975a, 313) draws the connection between confirmation holism and underdetermination:

> This holism thesis lends credence to the under-determination theses. If in the face of adverse observations we are free always to choose among various adequate modifications of our theory, then presumably all possible observations are insufficient to determine theory uniquely.[54]

Of course the quote from Quine above is not just addressing the question of the existence of empirically equivalent rivals, but also their significance for the question of evidential determination. We are now in a position to see this significance in light of an obvious problem with the argument for alluded to so far: the conclusion that evidence does not determine theory simply does not follow from the premise that there are empirically equivalent rivals. Some intermediate premise is needed to connect the two points. The likely intermediate premise is the second in the following argument:

> 1. For every theory, there is an empirically equivalent rival.
> 2. Empirically equivalent rivals are equally well-confirmed.
> 3. Therefore, for every theory, there is an equally well-confirmed rival.

The important question, then, is why we should accept Premise 2.

As it turns out, Quine seems to provide his rationale for the premise by a naturalistic appeal to a fact about scientific practice: that scientists rely on the hypothetico-deductive method to confirm their theories. Quine's clearest statement of the connection between hypothetico-deductivism and underdetermination is found in his essay "On Empirically Equivalent Systems of the World" (1975a, 313):

> If all observable events can be accounted for in one comprehensive scientific theory—one system of the world . . . —then we may expect that they can all be accounted for equally by another, conflicting system of the world. We may expect this because of how scientists work. For they do not rest with mere inductive generalizations of their observations: mere extrapolation to observable events from similar observed events.

[54] See also Quine (1981b: 9)

> Scientists invent hypotheses that talk of things beyond the reach of observation. The hypotheses are related to observation only by a kind of one-way implication; namely, the events we observe are what a belief in the hypotheses would have led us to expect. These observable consequences of the hypotheses do not, conversely, imply the hypotheses. Surely there are alternative hypothetical substructures that would surface in the observable ways.

Quine's allegation that observations can be "accounted for equally" by more than one theory "because of how scientists work" is as close to an explicit statement of the source of Premise 2 as we are likely to find.[55]

It should come as no surprise that this is his view about how scientists work, as he has made the point and its signification for confirmation elsewhere in his body of work. In "The Nature of Natural Knowledge" (2004d, 291) he says that the essence of the scientific method is hypothetico-deductivism:

> Science is a ponderous linguistic structure, fabricated of theoretical terms linked by fabricated hypotheses, and keyed to observable events here and there. Indirectly . . . the scientist predicts future observations on the basis of past ones; and he may revise the superstructure when the observations fail. It is no longer simple induction. It is the hypothetico-deductive method.

And in "Naturalism; or, Living within One's Means" (2004b, 276), he describes the implications of this method for a view of confirmation:

> [T]he most we can reasonably seek in support of an inventory and description of reality is testability of its observable consequences in the time-honored hypothetico-deductive way—whereof more anon.

[55] This passage should also put to rest any worries that the underdetermination thesis is *prima facie* non-naturalistic, because it uses the concept of "confirmation" (or justification, or support) in alleging that multiple theories are equally well-confirmed. There is no reason to think that the *concept* of confirmation or justification is unacceptable to the naturalist. Even if the concept has been the subject of many a conceptual analysis, naturalists are perfectly happy to explicate the concept for their own purposes. Clearly they are happy to take both deductive and hypothetico-deductive justification as paradigmatic forms of logical justification, and by assuming that there are no other forms of logical justification, argue that theories are equally logically justified if they have the same empirical consequences. To be sure, they can widen their naturalistic concept of justification to include pragmatic justification, but that is *as a result* of their conclusion about underdetermination of theory by evidence, i.e., the equal logical justification of empirically equivalent rivals. If there is something non-naturalistic about the underdetermination thesis, it is not something about its explicit formulation. Rather, later I will argue that the non-naturalistic assumptions are buried within the justification for Premise 2.

And: "[T]he deduction and checking of observation categoricals is the essence, surely, of the experimental method, the hypothetico-deductive method" (280).

Quine himself eventually distances himself from the idea that the global underdetermination thesis is epistemologically significant, or even coherent (1975a; 1976; 1979). But he does seem to take thesis as significant in "Two Dogmas of Empiricism" (1953b, 42–3) where he writes:

> The totality of our so-called knowledge or beliefs . . . is a man-made fabric which impinges on experience only along the edges. . . . [T]he total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience.

It is *this* brand of underdetermination that appears to inform his derivation of the indeterminacy of translation thesis and motivate his naturalized epistemology, both of which, as we have shown, are definitely epistemologically significant.

In any case, whatever Quine's ultimate view about underdetermination or its sources, we can take his view about the connection between the method of hypothesis and underdetermination as representative of a view that has been widely accepted throughout the history of science. Quine is not the first observer to conclude that an underdetermination problem seems to follow from facts about scientific practice related to the hypothetico-deductive method. In his survey of the history of the method of hypothesis, *Science and Hypothesis* (1981), Larry Laudan argues that the idea of *local* underdetermination is at least as old as the 3rd century B.C., when Apollonius first showed that motion of a planet on an epicycle was empirically equivalent to its motion on an eccentric. This equivalence, of course, played a central role in the later controversy over geocentric and heliocentric models of the solar system. Laudan notes that early religious commentators on the heliocentric hypothesis, such as Aquinas, Agostino Nifo, and Osiander (in his preface to Copernicus' *On the Revolutions*) all argued that because different hypotheses could imply the same astronomical consequences, the evidence by itself could not lead us to believe any of them as literally true—an early version of the argument for anti-realism from underdetermination, except as applied to only to a local domain.

Laudan argues in the first half of the 17th century, when science began to hypothesize more and more unobservable entities, a more global skeptical problem began to arise. Some scientists of the late 18th century (such as David Hartley and George LeSage) began to think of hypothetico-deduction as the exclusive source of scientific confirmation, and argued not even inductive generalization played the role originally defended by Newton and Reid. Underdetermination, once confined to the local domains to which hypothetico-deductivism was confined, such as astronomy, would now become more global. Of course none of these thinkers thought that all hypotheses were equally good, but they did think that they were equally well-confirmed by the evidence. As long as they regarded pragmatic criteria like simplicity as components of rationality, they did not become skeptics. Later developments would make this conclusion more difficult to avoid. And it was, perhaps, not a coincidence that Descartes, who championed the hypothetico-deductive method in science, also made an underdetermination-style skeptical argument concerning the evil demon (Okasha 2003).

Laudan emphasizes that these philosophers embraced the hypothetico-deductivist conception of confirmation because it was the conception—in contrast with inductivism—that best captured some of the most successful scientific theorizing of the day. They were unwilling, for example, to regard the wave theory of light as inadmissible simply because it posited unobservable entities (as Reid and other inductivists did), not when this theory met with such amazing predictive success (such as Fresnel and Poisson's prediction that a bright spot would appear under the shadow of a circular disk if the wave theory was true).These facts about successful scientific practice enriched their conception of confirmation—but paradoxically, the same conception also led them to regard the deliverances of science as subject to a form of skepticism.

We should, however, clarify just precisely how it is that hypothetico-deductivism implies Premise 2, that empirically equivalent rivals are equally well-confirmed. It is not enough that scientists sometimes find evidence in deduced empirical consequences. A stronger connection between hypothetico-deductive practice and the concept of confirmation is required. To say that two theories are empirically equivalent is to say that they have identical empirical content, i.e. the same empirical consequences. Stated loosely, hypothetico-deductivism is the idea that the empirical consequences of a hypothesis are a source of confirmation. But something stronger than that is needed for it to be the case that if two theories have the same empirical consequences, they therefore have the same degree of confirmation.

One way to support the idea that empirically equivalent rivals are *ipso facto* equally well-confirmed is to suppose that there is no way to confirm a given hypothesis H other than by observing *its* empirical consequences. In other words, observed empirical consequences of H are *uniquely* necessary for confirming H. After all, if other sources of confirmation are available, two empirically equivalent theories may have the same hypothetico-deductive confirmation, but differ in degrees of *other* forms of confirmation. In that case, empirical equivalence would not imply equal degrees of confirmation. Now as we shall see, it is highly implausible to think that a hypothesis can only be confirmed by *its* empirical consequences. Even the hypothetico-deductivist should acknowledge that a hypothesis might be confirmed in a straightforwardly deductive fashion, by deducing it *from* a better known proposition. But if that is the case and there is more to confirmation than observing a hypotheses' own empirical consequences, then more is needed to support Premise 2. *On its own*, Premise 2 does not seem to be true unless we also assume the implausible point we have just mentioned. Without that point, Premise 2 would be true only if the two theories also do not differ in their deductive confirmation, and it is an open question whether we can assume that. In the next section, we will consider arguments concerning whether or not the possibility of differential deductive confirmation is a problem for the underdetermination thesis. We will also consider whether there are forms of confirmation beyond deductive and hypothetico-deductive.

Premise 2 also presupposes a point that may seem much less controversial to many: that observing empirical consequences of a hypothesis *suffices* to confirm the hypothesis to some degree: once the empirical consequences of a hypothesis H are observed, this by itself is enough to provide H with some confirmation. So on this interpretation, if two hypotheses are empirically equivalent, they are both confirmed in some sense. If the mere ability to deduce an observable consequence from H did *not* suffice to confirm it, then the existence of equally confirmed rivals could not be assumed simply because of the construction of empirically equivalent rivals (through pure logic or through adjustment of auxiliaries): some empirical consequences might not be relevant to the confirmation of the hypothesis in question. So even if empirically equivalent rivals exist, and empirical consequences of hypotheses *often* confirm hypotheses, it would not follow that empirical consequences *always* confirm the hypotheses from which they are derived.

We now have a range of options before us concerning how to challenge the underdetermination thesis. We can challenge Premise 1 directly by challenging the idea that for any theory, there is always an empirically equivalent rival. And we can challenge Premise 2 in three separate ways: by challenging the idea that empirical equivalence implies an equivalence of deductive confirmation, by challenging the idea that hypothetico-deductive and deductive confirmation exhaust our forms of confirmation, and by challenging the idea that empirical consequences suffice to confirm. In the following section, I will argue that each of these premises can be called into question when we examine a wider context of scientific evidence that calls into question our original naturalistic sources of doubt.

Undermining underdetermination: the scientific roots in context

Premise 1: Are there always empirically equivalent rivals?

In an influential article, "Empirical Equivalence and Underdetermination," Larry Laudan and Jarrett Leplin (1991) challenge the idea that there *are* empirically equivalent rivals for every theory. They argue that theories with identical logical consequences are not necessarily identical in *empirical* consequences, because what it is to be an *empirical* consequence is relative to one's background theory, and varies from theory to theory. Changing auxiliary hypotheses in the manner envisioned by the confirmation holist may have the effect of changing a theory's class of empirical consequences, so it is not trivial that adjusting one theory when confronted will recalcitrant "observations" will result in another theory that is empirically equivalent. Much of the debate on underdetermination in the literature results from Laudan and Leplin's paper, and concerns this point about whether rivals generated by logical algorithm should count as empirically equivalent.[56] I should note, incidentally, that even if Laudan and Leplin's objection here misses the mark, it does show that what Bergstrom calls the "argument from logic" is irrelevant in the end to supporting Premise 1. The fact that we can find different hypotheses with the same logical consequences is unrelated to the question of underdetermination by *evidence,* because it does not address the question of *empirical* consequences. So clearly the argument from confirmation holism is the only serious argument for underdetermination.

John Norton (1994) raises another question about the existence of empirically equivalent rivals, this time more attuned to the naturalistic challenge of the argument from confirmation holism: "[O]ne rarely finds the scientist has such an embarrassment of riches. Typically, a scientist is pleased to find even one theory that is acceptable for a given body of evidence" (4). Of course it is possible that *some* actual empirically equivalent pairs may be found in the history of science. Alleged examples of these have included geocentric vs. heliocentric theories of the solar system, Newtonian mechanical theories differing in state of absolute rest, and Einsteinian relativity vs. Lorentz ether theory.[57] Of course not all of these remained empirically equivalent with the advance of theory, and even assuming that they offer actual examples of *local* underdetermination, it is harder to see that they threaten to establish any epistemologically significant, *global* underdetermination thesis.

---

[56] See Kukla (1993), Laudan and Leplin (1993), and Kukla (1996).

[57] Though even here, we note, with Laudan and Leplin (1991, 459) that it is interesting that all of the historical examples seem to stem from the relativity of motion. This strongly suggests that underdetermination effects can be isolated to a given domain of theory. This still leaves the question open as to how one is to choose among such rivals, but that open question would still seem to have little general epistemological significance.

In his later writings, Quine (1975a) became concerned that even if we find empirically equivalent rivals, they may not be the kind needed to generate a significant underdetermination thesis. This is because it is plausible that many cases that empirically equivalent theory formulations that are logically incompatible may be mere notational variants of each other, which can be rendered logically equivalent through an appropriate reconstrual of predicates. In the end Quine decided that there is no practical way for human beings to know that there are such reconstruals available, so rival theories may appear to be underdetermined. But he says it is also impossible to rule out the existence of such reconstruals, so whether theories are underdetermined amounts to an "open question." John Norton (2003b) makes a similar observation when he offers an interesting argument for why any empirically equivalent theory easy enough to construct in the space of a single journal article is probably just a notational variant of the original theory, not a genuine rival.

Whether or not there are empirically equivalent but logically incompatible *rivals* is an interesting question. If there are not, then not even the "weak" underdetermination thesis would be true; surely then the "strong" thesis would not be, either. However I think that the issues involved in resolving this debate relate little to dealing with the questions that originally gave rise to the underdetermination thesis. Here I have in mind more basic questions about the role of empirical consequences in confirmation in the first place. So we should ask about whether the existence of empirically equivalent rivals also implies that these rivals are equally well-confirmed, and question Premise 2. To do this, we should look to see if the wider context of scientific practice supports the hypothetico-deductivist conception of confirmation. We should ask: do scientists actually treat the deduction of an observed empirical consequence from a hypothesis as both a uniquely necessary and sufficient for confirming that hypothesis? I believe the answer is a strong "no," if counterexamples from the history of science presented by Laudan and Leplin (1991) and others are convincing.

Premise 2: Equal deductive confirmation?

Challenging the admittedly implausible claim that observing the empirical consequences of a hypothesis is the *only* way to confirm that hypothesis, Laudan and Leplin cite a historical example about the development of the theory of continental drift, call it T. This theory holds that every part of the earth has at one time occupied a different location, and implies the following hypotheses: H_1: The climate of each part of the earth has differed in the past; H_2: The current magnetic alignment of each part of the earth will differ with the alignment of older magnetic rocks. Laudan and Leplin then cite the fact that in the 1950s and 1960s, evidence *e* for H_2 was acquired in the normal hypothetico-deductive way (H_2 entailed *e*, and *e*—a set of facts about changing magnetic polarities in adjacent sections of the mid-Atlantic trench—was then observed). Interestingly, however, *e* thereby supported H_1, even though *e* was not an empirical consequence of H_1. The climate hypothesis H_1 was supported not because of *its* empirical consequences, but because it was a consequence of the theory of continental drift, T, which was hypothetico-deductively supported by H_2. Further historical examples of this kind of "indirect" confirmation are cited: the discovery of Brownian motion supported the atomic theory; the observation of the increase of mass with velocity supported relativistic principles; Thompson's cathode ray experiments supported hypotheses about electricity; Maxwell's observations of heat radiation supported the kinetic molecular theory; and the discover of heritable variation supported Darwin's hypothesis about the age of the earth (1991, 462). In none of these cases were the new discoveries empirical consequences of the hypotheses in question.[58]

The significance of these counterexamples is that while in some cases (as in *e*'s support for H_2), a hypothesis is supported by its empirical consequences, in other cases (as in *e*'s indirect support for H_1 via a theory which implies H_1), a hypothesis can be supported by what it derives *from*. This is

[58] Okasha (1997) objects that Laudan and Leplin do not specify the circumstances in which evidence flows upwards and those in which it flows downwards (253-4). That may be true, but I do not think it is incumbent on Laudan and Leplin to perform this task, particularly if, as naturalists, they are concerned with scientific practice. It seems to be sufficient that scientists sometimes regard one direction of flow as evidential, and sometimes the other—and that they meet with great success when they do this.

what I have referred to as "deductive confirmation" in the previous section. Laudan and Leplin call it "indirect confirmation" because the H_1 is supported not by its empirical consequences, but indirectly by virtue of deducibility from T, which is in turn supported by *its* empirical consequences (H_2, which is supported by its consequence, e). Earlier I had said that it seems unlikely that naturalists would discount this as an additional form of confirmation. But the existence of alternate forms of confirmation then implies that Premise 2—that empirically equivalent rivals are equally well-confirmed—cannot be supported unless we assume the further premise that empirically equivalent rivals are not unequal in deductive confirmation. We should now consider whether there is any good reason to assume this.

One objection to Laudan and Leplin's argument, raised by Bangu (2006), raises a point that might suggest why equal deductive confirmation of rival hypotheses might be assumed—thereby supporting Premise 2 of the argument for underdetermination. Suppose that our rival hypotheses H_a and H_b both imply the same empirical consequences, E. We shall consider whether they can be said to be equally well-confirmed in virtue of this empirical equivalence. Now one could use deductive confirmation to confirm H_a better than H_b, but only if there is a theory T_a (supported hypothetico-deductively by its own evidence *e*) which implies H_a but does not imply H_b. Bangu objects that the existence of such a T_a would not solve the underdetermination problem, because nothing rules out the existence of another theory, T_b, supported likewise by evidence *e*, the same evidence supporting T_a, which *does* imply H_b, but not H_a. So presumably whatever allows us to find the empirically equivalent pair, H_a and H_b also allows us to find a T_a and T_b which are empirically equivalent vis-à-vis evidence *e*. As long as there can be a T_a and T_b which imply different H's, we can assume, in effect, that H_a and H_b are equally well-confirmed just in virtue of being empirically equivalent. True, relative to T_a, H_a is better supported than H_b. But relative to T_b, H_a is better supported. Since there is no way to choose between T_a and T_b based on *e*, it seems that H_a and H_b are tied once again.

There are several responses to Bangu. First of all, I note that he admits that the existence of T_b is only a "logical possibility," not an inevitable fact of scientific practice. Yet I have emphasized that we should address ourselves to scientific practice to assess this question. If his response to Laudan and Leplin is to be a serious naturalistic source of underdetermination, there should be actual examples of T_a and T_b in the history of science—but none are presented. What's more, nothing about the logic of confirmation holism guarantees the existence of the higher-level theories needed to provide the necessary deductive confirmation for *any* hypotheses, to say nothing about multiple theories which differentially confirm rival hypotheses. At best, confirmation holism permits a way to find empirically equivalent hypotheses by permitting the arbitrary reshuffling of existing auxiliary assumptions. Once these have been reshuffled, it becomes deductively obvious what should be done with the original hypothesis. But finding new theories from which new hypotheses are to be deduced is quite another thing, involving not just the reshuffling of auxiliaries, but the discovery of new ones. Kepler's laws, for example, required Newton's discovery of calculus before they could be deductively confirmed by Newton's laws.

Second, Bangu considers an objection to his proposal which I believe he does not fully answer. Some would argue that that there could be no such thing as a T_b which is supported by the same evidence as supports T_a (that is, e), but which implies a different hypothesis than T_b (H_a rather than H_b).Bangu says it is trivial that two different theories, each entailing different hypotheses, can be supported by the same *e*. His example involves T_a, "All swans are white," and T_b, "All swans are white and all crows are black," each of which entail and are therefore supported by *e*, "All observed swans were white" (274).[59] There is a hypothesis, H_b, "All crows are black," which is entailed by T_b but not by T_a. Now although Bangu shows how it is possible for two different theories to entail different hypotheses while being supported by the same hypothetico-deductive evidence, he does not

[59] I note in passing that this T_b includes an irrelevant conjunct, which is not clearly supported by the evidence *e*. We will examine this issue later when discussing whether deriving an empirical consequence is sufficient for confirmation.

shown that the two different entailed hypotheses are *logically incompatible rivals*, which is the usual stipulation of an underdetermination problem. If H_b is "All crows are black," what H_a might be the different hypothesis supported by T_b? Presumably only "All swans are white." But "All swans are white" and "All crows are black" are not logically incompatible. So while the choice between H_a and H_b is underdetermined by hypothetico-deductive evidence, there is no reason to think a choice between them must be made! As we have already seen, two theories must be logically incompatible to be the kind of rivals that are interestingly underdetermined.

Presumably, however, it is a trivial matter to supply a pair of theories other than Bangu's which do imply incompatible hypotheses. Suppose, for example, that we selected as T_a, "All swans are white and All crows are black," and selected as T_b, "All swans are white and All crows are white." Each of these do indeed entail "All observed swans are white" and are therefore supported by it. But each can also be made to imply incompatible hypotheses: H_a, "All crows are black," and H_b, "All crows are white." Very well. But now there is a new problem. H_a and H_b achieve their incompatibility at the cost of no longer obviously entail the same evidence E, which is the original stipulation of the idea that they are empirically equivalent and therefore underdetermined by that evidence. Perhaps, however, this was a problem for Bangu's example to begin with, since it would also not have been obvious how "All swans are white" and "All crows are black" might have been entailed and supported by the same evidence. At least in that case, however, we were not dealing with hypotheses predicating opposite colors of the same species. Perhaps a story could be told about why white swans explain some piece of evidence, while black swans could also explain it. This overall objection—that Bangu's example does not meet all of the stipulations of the standard underdetermination problem—does not imply that there is no way to concoct the proper example. But it does show that it is not "trivially true" that two different theories, each entailing *logically incompatible* but *empirically equivalent* hypotheses, can nevertheless be supported by the same evidence. If it is not trivially true, then particularly because Bangu has produced no examples from actual scientific practice, there is no reason to think we can always assume that there always exist a T_a and T_b from which H_a and H_b might be deduced, rendering them equally confirmed and supporting Premise 2 (that empirically equivalent rivals are equally well-confirmed).

It is also noteworthy that Bangu assumes that the only source of support for T_a and T_b is itself of a hypothetico-deductive nature. This is a fair assumption, since Quine and other major parties to the discussion assume this. “Hypotheses,” Quine says, are “related to observation *only* by a kind of one-way implication”—from hypothesis to observation, not vice versa. But it is worth thinking about whether we should assume this. If there are other ways for our theoretical knowledge to interface with experience, this would mean that there could be ways aside from hypothetico-deduction in which hypotheses can be empirically supported. If there is, then there is another way in which Premise 2 might fail: two hypotheses with identical empirical consequences might still be differently confirmed.

Premise 2: Alternative sources of empirical confirmation?

Laudan and Leplin themselves entertain one additional source of independent empirical support: analogy. They identify important examples of reasoning by analogy in the history of science. They point out how Maxwell drew an analogy between an enclosed gas and an enclosed set of colliding particles, supporting the molecular theory of gases, and how Einstein drew an analogy between the entropy of certain kinds of radiation with that of ideal gases, supporting a quantum view of radiation (1991, 464-465). Surely there are others. If full-fledged theories can be confirmed by analogy, then surely perhaps individual hypotheses might be, as well.

Of course analogy is not always an independent source of confirmation. An analogy drawn with the behavior of ideal gases would be only as strong as our knowledge of these gases, which undoubtedly requires much independent confirmation of its own. Consider, however, Maxwell’s analogy between gas molecules and colliding macroscopic particles. Confirming our knowledge of the properties of colliding particles more clearly involves direct observation, and plausibly, inductive generalization. Understanding the way analogy can draw on inductive generalization is particularly important for understanding a common criticism of inductive generalization made by exclusive advocates of hypothetico-deductivism: that inductive generalization cannot *directly* generate hypotheses about unobservable entities, like light waves or atoms. This fact has been used by some hypothetico-deductivists to argue that, contrary to inductivists, the inductive generation of hypotheses is irrelevant to their confirmation. Laudan (1981) reports that some hypothetico-deductivists (here I have in mind Auguste Comte and possibly William Whewell) went so far as to suggest that inductive generalization has no role to play at all—that *only* hypothetico-deduction confirms hypotheses. But even if it is true that there can be no direct inductive generalizations about unobservables, it is possible that many hypotheses supported by analogies to entities we *can* generalize about.

Consider another example concerning unobservable entities: even if we cannot observe and generalize about atomic nuclei, we *can* observe and generalize about larger massive objects. When Rutherford bombarded gold foil with alpha particles, and found that the majority passed straight through while a few were deflected at drastic angles, he is reported to have compared the discovery to naval shells being deflected through tissue paper. Presumably only tiny but massive entities scattered throughout tissue paper could account for naval shell deflection, so Rutherford was led to conclude that there must be proportionally similar objects embedded in the gold foil. Famously he concluded that the bulk of atomic mass was concentrated in a nucleus; much of the rest of the atom was empty space. Now of course Rutherford had probably never made any special study of tissue paper and naval shells, but he doubtless had commonsense *generalized* knowledge about the properties of macroscropic objects such as: massive bodies ordinarily pass through thin sheets of flimsy material. And granted, he probably never bothered to explicitly consider or write down such an obvious piece of general knowledge, but it is difficult to imagine that it was not an important element of background knowledge leading to the generation of his original nuclear hypothesis (and as we shall see in the final section of the chapter, there is no good reason to think that every justification must be written down).

It is widely recognized that *arguments* by analogy can play a justificatory role, but a space needs to be made for the possibility of implicit analogical justification deriving from background inductive generalization. Of course the support analogy provides is not complete support: once the hypothesis is proposed, rigorous hypothetico-deductive testing along various parameters would be necessary. But allowing that background analogies can offer support would also allow us to see hypothesis formation as more than merely psychological, but also as logical. Premise 2 would then suffer: identical empirical consequences would not render two hypotheses equally well-confirmed.

All of this, of course, depends on the justificatory power of inductive generalization, and here Hume's problems loom. In the next major section of this chapter, I will turn directly to Hume's problems, and possible scientific solutions to them. But for the moment it is worth addressing some preliminary worries about how inductive generalization might serve as an independent support for hypotheses. Sarkar (2000) considers just this question, in relation to assessing Leplin's (1997) critique of the underdetermination thesis. Sarkar (2000, 190) notes a tension between the possibility that auxiliaries might be confirmed inductively, and Leplin's view that "brute inductive generalization" is itself not justifiable in the absence of some nomic or causal explanation of the generalization (1997, 24, 112–115). And, since nomic/causal explanations would need to be justified hypothetico-deductively themselves, inductive generalization could never be a source of independent support, at least on Leplin's own premises.

Sarkar is correct that there is a tension here, but I believe that it stems from Leplin's failure to articulate his own views precisely. In an interesting passage, Leplin notes that inductive generalizations may be supported by "explanatory" connections which themselves do not depend on "antecedent inductions." He writes:

> There is more to causal agency than the observed concomitances that Hume thought exhausted our perceptual access to agency. The idea of agency is supported by elementary impressions of agency, as when we experience effort or pressure. Such experience is the primitive basis of the idea of force, which is undefined, or only implicitly defined, in classical physics. A chalk mark made on the blackboard is not simply the concomitance of the chalk's motion with the appearance of the mark. One

> is directly aware of *making* the mark, of *imposing* that chalk upon the board…[T]he possibility of an experience of agency unmediated by generalization from instances supports, at least, an epistemic parity between inducing and explaining. (1997, 117–8)

Here Leplin is appealing to a kind of direct realist justification of generalization that is highly compatible with the account of the empirically given that I will propose at the end of this chapter, an account which relies in part on the idea that causal relations might be directly perceived, thereby allowing for the possibility of inductivist foundationalism. If there is any causal relation directly perceived, Leplin is surely correct that one's own agency may be the most fundamental, whereupon other forms of causality are attributed to nature by analogy. But in contrasting this view of "explanation" with inductive generalization, Leplin is assuming that induction begins with Humean regularities, stated in the form of singular propositions. One need not accept this if one thinks that even singular propositions, insofar as they involve the use of concepts, already involve generalizations. This is, in fact, the position of Norton (2003a) and Lange (2004), to which I shall turn at the end of the chapter. If inductive generalization itself is non-Humean, and is *constituted* by this direct realist "explanation," it is not the case that inductive generalization needs some separate explanatory justification.

In the last few paragraphs, I have spoken about forms of confirmation beyond hypothetico-deductive confirmation, in order to directly challenge the supposition that confirmation is ultimately hypothetico-deductive, and therefore to show that equality of empirical consequences does not imply equal confirmation. But there may still be some question as to how the use of alternative sources of confirmation works in practice, especially as regards using these sources to independently confirm or disconfirm the auxiliary hypotheses that are otherwise subject to the argument from confirmation holism. The idea that the independent confirmation of auxiliaries is a solution to the underdetermination problem is shared by Leplin (1997, 152–8) and also by Elliot Sober (1999). But it might be objected that expecting all of the relevant auxiliaries to be fixed independently is unrealistic, and therefore not a plausible solution to underdetermination.

We can only grapple with this question by breaking down types of auxiliaries into categories. The simplest type is discussed by Sober: assumptions about the reliability of particular scientific procedures. He considers the assumption that a pregnancy test is reliable: typically, he notes, the outcome of a pregnancy test does not alter one's confidence in the reliability of the test. Very different tests would be needed to judge this matter. Or consider my example: Galileo's telescope. Galileo's confidence that the scope allows him to see mountains on the moon is clearly based on prior calibration of the instrument with already-known terrestrial observations. These kinds of auxiliaries can clearly be supported independently by straight induction, together with *some* understanding of the causal mechanism involved. True, understanding the optics of the telescope is based on some prior theory, but it is theory that is closer to direct inductive generalizations than the theories he uses his telescope to support. Galileo did not need a wave theory of light, for example, to understand the relevant optics of the telescope, only the ray optics of the lenses and how they combine to form a certain image.

A second type of typical auxiliary is the *ceteris paribus* clause. Consider the relevance of such clauses in the testing of Newton's theory of gravitation: it was originally used to predict the orbits of both Uranus and Mercury to be different than they were actually observed to be. The original (false) Newtonian prediction about planetary orbits could be seen as following with the assistance of the eventually discarded auxiliary assumption, "There are no trans-Uranian planets." So the original prediction assumed that all things were equal, that there weren't interfering planets. But the prediction was false, so scientists called into question their assumption that all things were equal. A new auxiliary assumption was considered: that there *was* a trans-Uranian planet causing the perturbations. When such a planet was actually discovered (Neptune), the new auxiliary was independently confirmed, and Newtonian theory was saved from refutation. Before it was independently confirmed, it was merely a working hypothesis, not actual grounds for retaining Newtonian theory.

Of course the outcome was quite different in the case of Mercury. In that case, the original auxiliary (that there was no intra-Mercurial matter) was retained, while Newtonian theory was jettisoned in favor of general relativity. The difference between these two cases is often cited as evidence that the choice of hypothesis/auxiliary combination is pragmatic, even a result of social factors (such as the pressure to adopt a mathematically elegant new theory rather than retain the increasingly difficult to defend old theory). But there are conceivably good independent reasons that that the original auxiliary was retained in the case of Mercury, while it was discarded in the case of Uranus. It is important, for example, that whereas a trans-Uranian planet was actually found, intra-Mercurial never was, even though the latter, being confined to a smaller space closer to the Earth, should have been easier to detect, even by scientists of the day. That it should have been easier was itself a further auxiliary hypothesis, much like our assumption that we would see a large elephant in the room if it were here: since we do not see it, we conclude there is no such thing. Both of these, I think, derive from the general background knowledge that things are easier to find in smaller spaces. The same auxiliary would not have been applicable in the case of Neptune—more like a needle in a haystack than an elephant in a room—so there would have been reason to keep looking for a longer time without officially discarding Newtonian theory.

Other auxiliaries concerning unobservables are not so easily independently confirmed. An interesting case, also discussed by Leplin (1982), is Pauli's postulation of the neutrino to account for beta decay in a manner consistent with the conservation of energy. In this case, neutrinos would be in principle unobservable (unlike distant planets), so there is a serious question about how any auxiliary hypothesis used to predict facts about beta decay could be supported independently, rather than simply giving up the conservation of energy principle. Now I think that the neutrino hypothesis itself is formed by analogy to other observable cases in which conservation laws are preserved pending the discovery of some new interacting agent (for example, Lavoisier's early conservation of mass principles). Apart from that, the hypothesis was also eventually supported by *independent* hypothetico-deductive evidence.[60] Perhaps before this independent support was offered, energy conservation of beta decay was temporarily called into question in the intervening period, rather than accepted.[61]

So, yes, hypothetico-deduction is a crucial source of scientific confirmation. But it is not the only source or the *ultimate* source. As long as there are other sources, there are independent ways to confirm both auxiliaries and stand-alone hypotheses, and therefore ways to confirm empirically equivalent pairs differentially.

Premise 2: Do empirical consequences always confirm?

Once we accept scientific practice as a standard for evaluating conceptions of confirmation, I think it is also easier to show how observing an empirical consequence of a hypothesis does not always suffice to confirm it. If this is true, then simply from the fact that we might find a hypothesis that is empirically equivalent with another, we cannot infer that it is equally well-confirmed, because we cannot conclude that it is even confirmed at all.

To show this, we need only give examples of empirical consequences that no scientist would take as confirming any hypothesis. One traditional objection to hypothetico-deductivism on this front is the "problem of irrelevant conjunction," which notes that if observing an empirical consequence suffices to confirm a hypothesis, and if *e* is a logical consequence of H, then the conjunction of H with any arbitrary X is also confirmed by *e*, since anything that is a logical consequence of H is also a logical consequence of H&X. Of course no scientist would take any arbitrary conjunction as confirmed by the consequences of only one of the conjuncts.

There are more realistic examples of hypotheses that are not confirmed by every one of their logical consequences. Laudan and Leplin note that observing young men entering puberty after reading scripture does not confirm the strange hypothesis that scripture-reading induces puberty, and

---

[60] In 1956 the important Cowan-Reines experiment detected neutrons and positrons that should have been produced by the interaction of free neutrinos with protons.

[61] For more on how theories involving auxiliaries about unobservables might overcome underdetermination problems, see Michela Massimi's (2004) application of the method of demonstrative induction, rather than hypothetico-deduction, to aspects of quantum theory.

that the point of controlled experiments is precisely to rule out those empirical consequences which are not confirmatory (1991, 465–6). Even for theories not open to experimental testing, serious scientists do not consider theories with identical observational consequences equally well-confirmed. A good example from John Norton (2003b, 15) is the empirical equivalence between standard geology and young Earth creationism, as regards the fossil record: provided that the young Earth creationists tell the right story about how God put the fossils there to test our faith, the evidence we observe may be exactly what the latter theory predicts. Yet scientists do not take the second view seriously.

Bayesian epistemologists may argue that the objection above confuses two senses of "confirmation." They might say that while it is true that scientists do not take these arbitrary hypotheses as *absolutely* confirmed (i.e., having a probability above a certain threshold, warranting acceptability), this does not mean they do not take them as *incrementally* confirmed (i.e., having probabilities given the evidence which are *higher* than the prior probability of the hypotheses).[62] On this view, observing an empirical consequences does raise the probability of a hypothesis—even if only slightly—and so observing empirical consequences always suffices to confirm hypotheses incrementally. This is essentially the solution Maher (2004) presents for the problem of irrelevant conjunctions, which is analogous to the current problem: both objections suggest that hypothetico-deductivism is committed to the confirmation of strange hypotheses (either intrinsically strange ones, or ones paired with strange conjuncts). Maher suggests that an irrelevant conjunction *is* confirmed by *e*, though it is only incrementally confirmed. If scientists do not accept that these kinds of hypotheses are confirmed, then they are simply talking about a different kind of confirmation than incremental confirmation.

I am of two minds about this response. On the one hand, even if it is correct that hypotheses are always confirmed in this incremental sense by the observation of their empirical consequences, this does not yield the *kind* of sufficiency requirement needed by the underdetermination thesis. Just

[62] See Maher (2006) for the distinction.

because two hypothesis, one respectably scientific, the other a creation of fantasy, are both incrementally confirmed to some degree does not mean that they are incrementally confirmed to the *same* degree. Many measures of incremental confirmation are defined in terms of the prior unconditioned probabilities of the hypotheses, and if these are different, their probability on the evidence will be different as well. Other measures in terms of the probability of the evidence on the negation of the hypothesis have a similar outcome. So I could readily concede this kind of sufficiency clause, but show that it defeats Premise 2, and thereby underdetermination.

At the same time, I think there are reasons to question the claim that hypotheses are always confirmed incrementally by observation of their empirical consequences. In Bayesian epistemology, a hypothesis is incrementally confirmed on the assumption that the probability of the hypothesis given our background knowledge is as follows: $0 < P(H/D) < 1$. It has been traditionally assumed that as long as H (and D) involve no outright contradictions, $P(H/D)$ is indeed greater than 0. Perhaps it is this assumption that must be rejected. Perhaps to have some degree of probability given background knowledge, it is not enough that no contradictions obtain, that the hypothesis be conceivable. Perhaps there needs to be some special reason to think a hypothesis is possibly true—the kind of special reason often suggested by analogies or inductive evidence. Indeed it is not at all clear why *deductive* logical possibility should be a guide to empirical *inductive* probability.

In general, I think there is good reason to believe that defining confirmation in terms of probability is also putting the cart before the horse. I think "confirmation" is a more basic concept than "probability." We say that a proposition has a high-probability if it has been well-confirmed. Probability is a measure of the degree of confirmation; it isn't that confirmation is a measure of the degree of probability-raising. This is consistent with my suggestion above that some evidence—some confirmation—is needed before we can say that a hypothesis has even a tiny degree of probability. I doubt that "probability" is a primitive concept. Our idea of degrees of probability derives from the fact that we know that we can have different amounts of evidence for the same conclusion as we collect new facts. We come to this realization only after we have collected all of the evidence we think possible, and then realize that we didn't always have this much evidence. Probability is also clearly not primitive in that it is a modal modifier: a claim can be probably *true,* or possibly *true*, and we see these two different modalities of truth as falling along a continuum. The etymology of "probability" is also suggestive: probability derives from *provability,* and different degrees of probability do indeed reflect degrees to which a claim has been proved. Some facts about the history of science also support this. Philosophers and scientists debated theories of confirmation well before they had theories of probability. They also did not regard irrelevant conjunctions as confirmed. Now it is possible that when probability theory entered the field, and began to inform scientists' decision-making, there was a Kuhnian paradigm shift, and the old concept of confirmation was simply incommensurable with the new concept. But I doubt this. I have already given some reason to think that forms of confirmation that at first appear to be discontinuous (such as inductive generalization, analogy, and hypothetico-deduction) may in fact be related as fundamentals to derivatives. I suspect something like this is also true about the relation between confirmation theory and probability theory, though this has not yet been worked out (Norton will suggest some leads in our next section).

Of course I do not even need to defend against incremental confirmation to undermine the underdetermination thesis, and that is my primary task. Even if arbitrary hypotheses are incrementally confirmed, rival hypotheses are not necessarily equally incrementally confirmed, so there is no underdetermination problem that results from this formulation of the sufficiency point. And even if the sufficiency requirement cannot be refuted, the other attacks on Premise 2 (and on Premise 1) are still relevant. In any case, looking at the wider context of scientific practice shows us how to dissolve a skeptical problem that at first appeared to arise from scientific practice. We have used Quine's own putative strategy, of urging the skeptic not to "overreact," to disavow underdetermination, one of Quine's most basic skeptical theses.

Concluding notes on pragmatism and confirmation

In the above, I have surveyed some aspects of scientific practice that I take to be relevant in showing the impoverishment of the idea that all and only empirical consequences confirm. One question worth asking, however, is why I choose to admit the examples of scientific practice that I do. Some might object that the only criterion I could appeal to is a pragmatic one: scientists regard this and that as a source of confirmation because achieves successful prediction and explanation, etc. If that is in fact my criterion, then there is something disingenuous here to Quine. I noted at the very beginning that I would distinguish two senses of "confirm": a pragmatic and a non-pragmatic one. I acknowledged that Quine held that many theories are better *pragmatically* confirmed than empirically equivalent rivals, but wanted to decide if he was correct that theories were not determined by *logically* confirming evidence. If now I claim to show that other sources of confirmation are available—but only because of examples that have been chosen because of their pragmatic significance, then this might be taken as sleight of hand. I would be arguing that additional sources of logical confirmation are possible because pragmatic confirmation is possible, which would neglect the very distinction I attempted to begin with.

It is true that we could understand the choice of scientific examples as stemming from pragmatic considerations. But we do not need to. At minimum, I could say that I pick examples of scientific theories that are widely accepted by my audience for a variety of reasons. The question is *why* they are accepted. Some might accept them because of the pragmatic value they exhibit, others not. But in the course of presenting some of these theories, the reason these theories are accepted is not pragmatic at all. They may be based ultimately on analogy and induction, which are evidence-based forms of confirmation. Now it might be thought that this is a question-begging strategy. Whether or not evidence-based confirmation is possible is the very thing in question, so I should not simply propose that it is possible and show where more of it might be. But this charge is to misunderstand that logical confirmation is being question by a *reductio ad absurdum*, which permits us to accept certain premises as true in the interim, only to see that they yield unacceptable consequences and must be rejected. In the above, I have simply provisionally assumed that more forms of evidence are confirmatory than the hypothetico-deductivist is originally willing to assume. In the same way, the underdetermination skeptic would assume a smaller range of forms of confirmation in order to test whether such confirmation helps to determine theory choice. (Quine, for example, would assume that observation sentences can be accepted as true; the question is whether they yield the truth of theoretical sentences.) In this way, I help to dissolve the skeptic's *reduction* by showing that the absurdities result only from impoverished original assumptions.

Of course there could still be independent reasons for doubting even the evidential relevance of induction, for instance, and insisting that hypothetico-deduction is the only possible form of confirmation. Interestingly, however, the deeper the skepticism becomes, the harder it is to see what kind of scientific problems might have prompted these doubts. It therefore becomes increasingly plausible that purely philosophical assumptions are at work. Laudan and Leplin suggest, for instance, that the 20th century commitment to hypothetico-deductivism derives from central problems in logical positivism. They note that the positivists' project was motivated by the wish to dismiss favorite targets (such as metaphysics, religion, or ethics) by proposing criteria according to which these targets counted as meaningless. Their goal was primarily epistemological: to characterize a certain kind of belief as irrational. But their means was semantic: to devise formal criteria "demarcating" the meaningful from the meaningless. The epistemic goal was then reached through the semantic injunction that the meaningless not be entertained. As a form of deductivism, hypothetico-deductivism helped flesh out this semantic injunction (1991, 466–8). First, positivists sought to define meaningfulness in terms of deducibility *from* observation statements ("strong verifiability"). As Soames (2005) outlines, it was only because this standard implied that too much ordinary scientific discourse would be meaningless that hypothetico-deductivism ("weak verifiability") was proposed: much more would count as meaningful on this latter view because it was far easier to find particular empirical consequences of scientific theories than it was to translate them wholly into sensory language.

With all of this in mind, it is worth wondering: *why* fight an epistemic war with logico-semantic proxies? Why not engage the enemies directly, by defining standards of epistemic justification in terms of something else? Why not, for example, define an *inductivist* standard of epistemic justification, rather than so many deductivist varieties? This is a question Quine answered explicitly, in his presentation and critique of Carnap in "Epistemology Naturalized." Carnap's project of offering translations of scientific statements in terms of observation language was what Quine called a "conceptual project," specifying a logico-semantic criteria of cognitive value, offered for the sake of a wider "doctrinal" goal, i.e., legitimating scientific discourse by showing how its language could be eliminated in favor of respectable logical and set theoretical terms. Of course empiricist doctrinal goals *had* been more ambitious. Rather than simply *exculpating* scientific discourse—showing that it was at least not guilty of any major crimes concerning meaning—empiricist tradition had once sought to *justify* scientific knowledge. But Quine believes this effort had originally met its match when it faced Hume's challenge: generalizations and singular statements about the future proved to be impossible to deduce from experience. Carnap's project never advanced much beyond this "Humean predicament" because these same statements remained untranslatable and Carnap looked only for ways of stating their consequences. So, the reason that empiricists fought their battle with logico-semantical proxies was their prior conclusion that a full-frontal epistemological assault was futile. Quine, famously, extended the devastating implications of the Humean predicament by urging that any further attempt even at merely exculpatory translation, had to fail—because of his derivation of the indeterminacy thesis from confirmation holism and/or the underdetermination thesis.

But now we are in a unique position to reassess the original Humean predicament. Having dispensed with underdetermination-style skepticism by resorting to the free (non-question-begging) use of science, could we apply a similar strategy to more traditional anti-inductivist skepticism? To see whether we could, we first need to survey some possible theories of induction. Having examined these, we can revisit the standard Humean doubts to see whether they, too, arise as scientific doubts. If they do, we may be able to make further free use of science to block yet another *reductio*. But, if it turns out that Humean doubts themselves make some questionable purely philosophic assumptions, then Quine is wrong that all skeptical doubts arise in a scientific context. If he is wrong about that, it is devastating to his naturalism, because it will mean that Quinean naturalized epistemology is motivated in part by purely philosophic concerns—not scientific problems.

Actually, before moving on to examine any of these questions, I want to draw a line between what I have accomplished above and what I suggest below. While it is true that a full solution to the underdetermination problem requires showing that there are evidence-based sources of confirmation apart from hypothetico-deduction, it is fairly plausible that debates about whether these sources are available will be more purely philosophical than scientific. So I think that once I have shown that one of Quinean naturalism's main motivations—the underdetermination thesis—rests on more basic Humean questions about induction, then regardless of how those questions are to be answered, much of the critique of Quinean naturalism is already complete. Even if there are unchallengeable reasons to be skeptical about induction, the fact that these are likely to be philosophic rather than scientific reasons already undermines the naturalistic claim that skeptical doubts are prompted by science—and hence the claim that philosophy itself is continuous with science. For this reason, most of what follows should be considered as merely suggested alternatives to Humean doubts. This is not a dissertation on induction or perception, and I cannot hope to defend the views I suggest below in much detail.

A scientific solution to Humean doubts?

Since we are curious about the possibility of confronting Humean doubts head-on, we should look for accounts of induction that a) do not rely on any logico-semantic criteria, and b) that draw inspiration from the history of science. An ideal candidate for such an account was proposed by John D. Norton in his recent, fascinating paper, "A Material Theory of Induction" (2003).

Norton argues that attempts to understand inductive confirmation in formal terms face an inevitable tension between the universality of their scope and their success in describing inductive relevance. An example of this tension is seen in the view that induction proceeds by simple enumeration. Norton gives an example from Mill, noting the difference between the inference from "Some samples of element bismuth melt at 271°C" to "All samples of the element bismuth melt at 271°C" on the one hand, and the inference from "Some samples of wax melt at 91°C" to "All samples of wax melt at 91°C" on the other. One way to state a universal formula for induction by simple enumeration might be to say that we can always infer "All S is P" from "Some S is P." But a formality like this does not explain what makes the inference about bismuth superior to the inference about wax.

In order to state standards that underwrite successful inductions, Norton thinks that a criterion must not be formal or universal. This is why he thinks "All induction is local," i.e., all inductive inferences are licensed by knowledge of particular matters of fact, relevant within particular domains, rather than by a formal scheme applicable universally. His positive case results from surveying various theories of induction, and concluding that each theory usually does describe one aspect of inductive practice, but that the successful use of this practice depends on specific matters of fact related to the subject matter. For example, he notes that the usual way of explaining the superiority of the bismuth inference over the wax inference is comparable to Quine's account of the superiority of "green" vs. "grue" inferences (1969b): the first is regarded as a natural kind concept whereas the second is not. Whereas wax is a more variegated family of substances, bismuth is a chemical element, bringing with it the expectation of more uniform associated characteristics. But whether a particular concept designates a natural kind is itself a question depending on our local scientific (e.g. chemical) theories. To embellish Norton's point: "wax" may have good inductive potential within certain domains: from the fact that some wax melts it is reasonable to conclude that all or even much of it does, without specifying a melting point. But only special scientific theory gives us confidence in generalizing about something as specific as a melting *point*, e.g. the specific chemical knowledge contained in our concept of "bismuth." So the inductive relevance of simple enumeration—if there is any—is not a merely formal matter, but a matter of specific facts about the concepts involved in enumerative inductions.

Norton goes on to make parallel points about other theories of induction, especially the hypothetico-deductivist and probabilist views. We have already encountered the pitfalls of the formalist approach to hypothetico-deductivism. Understood purely formally, this view sanctions "promiscuous" inductions, in such a way that *any* hypothesis, no matter how arbitrarily selected, would count as incrementally confirmed just in case some actually observed empirical consequences are deducible from it. The problem of "irrelevant conjunctions" is one such example. If my development above is correct, and this view of confirmation is unacceptable by reference to actual scientific practice, then to the extent that empirical consequences *are* ever confirmatory, some constraint must be placed on which conjunctions would count as confirmed. Norton points out that a judgment about which conjuncts would count as "relevant" would, again, be a matter of local fact as judged by our background knowledge. Similar constraints could prevent the consideration of entirely arbitrary hypotheses, such as the one about scripture reading causing puberty.

Norton then surveys a variety of theories of confirmation intended to be constraints on the original hypothetico-deductive model. Exclusionary accounts, for example, supplement hypothetico-deductive confirmation with some demonstration that the *falsity* of the hypothesis would make it unlikely to observe the empirical consequences. Demonstrating this, Norton thinks, would depend on facts of the matter concerning, for example, what counts as the randomization of the sample in a particular controlled-experimental setup. Simplicity accounts try to show that only "simple" hypotheses are confirmed by empirical consequences. But Norton thinks that judgments of simplicity (as in the stock example of curve fitting) are essentially judgments about whether or not to consider some causal factor as operative in the particular domain of facts (as when deciding to raise the power of the polynomial equation fitting the data points). The inference to the best explanation version of hypothetico-deductivism is similar: empirical consequences confirm a hypothesis only if the truth of the hypothesis would also help explain them. But what it is to explain in a particular domain (e.g. physical vs. biological vs. social), and what factors count as relevant to explanation (e.g. which types of causes, if explanation is about displaying causes), are judgments concerning local matters of fact. Finally, much the same is true about judgments of reliability involved in reliabilist views.

Norton thinks that a final class of confirmation theories—probabilist ones—are the most successful in making a case for a universal inductive scheme. Most of his discussion concerns subjective Bayesianism, which he critiques for reasons unimportant here. His most important comments concern Bayesianism as such. However probability is to be understood (as degree of belief or some objective relation to evidence), Norton urges that the calculation of probability conditional on evidence is determined by "factual properties of the relevant system." He thinks this holds even when we try to determine if the hypothesis H, "All swans are white," is confirmed by the evidence e that some particular swan is white. In this case, H entails e, and $P(e|H)=1$, but we must assess $P(e|{\sim}H)$, to know the probability of finding white swans if not all swans are white, and this depends on the prior probability $P(e\&{\sim}H)/P({\sim}H)$. Presumably he thinks this and other priors could only be assessed by reference to "factual properties of the relevant system." [63] Presumably he would make the same point about the dependence of $P(H|e)$ on $P(e|H)$, where $P(e|H)$ is to be determined by our knowledge of how the world would be if H were true of it.

Having surveyed various views of induction, to show that each has its place depending on the domain of facts in question, Norton presents the most remarkable part of his paper: a discussion of how a material theory of induction avoids Hume's problem of induction. Famously, Hume's problem

[63] It is, of course, not uncontroversial to think that prior probabilities must be factual in some way. Maher (2006) denies this, holding that they must be understood in a purely logical, *a priori* manner. I address these concerns in an online appendix at http://www.benbayer.com/blog/archives/2007/03/do_probability.html.

claims that there can be no justification of induction. Such a justification would be either deductive or inductive. If the justification is deductive, it fails to provide ampliative inferences beyond observation reports, as universal generalizations cannot be deduced from singular statements. If the justification is inductive, it either generates circularity or an infinite regress: circularity, because relying on the past success of induction to infer its future success would clearly be using induction; an infinite regress, because citing some principle like "the uniformity of nature" to bolster our inductions would then require some further principle as justification (how do we know nature is uniform, and in the relevant respects?). Norton observes that many of these problems might also befall a material theory of induction. Just as a deductive justification of induction would fail to ground its special inductive character, so some *a priori* postulate of a material fact of the matter would likewise sunder material induction. Just as the justification for a formal inductive schema would be circular if it applied the very same schema, a justification of a material induction would be circular if it used the same material fact referenced in its conclusion. The chief difference between formal and material views of induction concerns the last possibility, regress.

A formal schema of inductive inference could justify itself by reference to some different inductive schema (like the uniformity of nature), and further meta-justifications of this schema would likely be more artificial and abstract. A material induction faces no such problem. If the conclusion of one material induction is licensed by some different *particular* fact, the conclusion of an earlier material induction, there is a regress involved, but there is no reason to think it must be infinite. Rather than becoming more "fanciful" and abstract, the regress of material induction could gradually become more familiar and concrete. Norton (2003a, 668) concludes:

> The regress described here…merely describes the routine inductive explorations in science. Facts are inductively grounded on other facts; and those in yet other facts; and so on. As we trace back the justifications of justifications of inductions, we are simply engaged in the repeated exercise of displaying the reasons for why we believe this or that fact within our sciences.

Consider Norton's example, "All samples of the element bismuth melt at 271°C." This generalization was not supported by a simple formal rule, such as the general relevance of inferring "All S is P" from "Some S is P" (or "All observed S is P," etc.). What gives us confidence in this generalization, unlike the similar generalization about the melting point of wax is the specific knowledge that bismuth is a separate chemical element, plus our general knowledge about the uniform nature of chemical elements. Bismuth's status as an element is justified by isolating it from its oxides (bismite, bismuthinite or bismutite), and from the subsequent inability to isolate it further by chemical means. The bulk of our confidence in the induction comes, of course, from subsuming bismuth under our concept of "element," which is informed by quantum theory, atomic theory, and early modern chemistry, each of which were founded on a series of observations and experiments. We are confident in our bismuth induction because we are confident that quantum theory explains the properties of elements by reference to atomic structure. We are confident in the existence of atomic structure because of the work of by Thomson, Rutherford, Dalton, etc. We are confident that this work relates to standard chemical methods of isolating elements because of Lavoisier's early work with the most basic of elements (hydrogen, oxygen, etc.), establishing chemistry as a discipline. Each of these theories is in turn supported by a series of experiments on particular substances. As our justifications take us back into the historical foundations of chemistry, the facts to which we appeal become more and more concrete, like Dalton's observations about the combination of elements in multiple, integral proportions, Lavoisier's observations about the conservation of mass of the reactants in combustion, etc. In the material theory of induction, the context of discovery and the context of justification begin to intertwine.

It is noteworthy that with the material theory, a source of ambiguity that exacerbated Hume's problem is exposed: that between the attempt to justify *induction* and the attempt to justify *particular inductions*. Treating inductive justification as the first leads to a formalist conception of induction, and with this are associated the usual Humean problems. But treating inductive justification as the need to justify particular *inductions*, i.e. particular inferences, not only offers the hope of avoiding the infinite regress, but also the promise of doing so by attending seriously to actual scientific practice—just what our anti-skeptical strategy recommended.

Norton does face one serious problem. While the conception of material induction does at least make it more plausible that there need be no infinite regress of justification, it is not clear how the regress would terminate. Norton himself is not entirely sure. He suggests that it is possible that the regress could terminate in beliefs only admitting of circular justification, but they could also end in "brute facts of experience" in no further need of justification. Norton takes the uncertainty here as good news: as long as the foundationalist option is open, there is no reason to think that Hume has offered a knock-down argument against the possibility of justifying inductions. I think Norton is right about Hume, but it is worth considering whether latter-day empiricists may have gone any further. The idea that termini in "brute facts" of experience could provide epistemic justification ("empirical foundationalism") is precisely what is denied by coherentist empiricists like Wilfred Sellars (1963), and Donald Davidson (1983), both of whom agree that treating the "empirically given" as a source of justification is a "myth." The Sellars-Davidson objections to the given usually involve the following two points: 1) experience at best causes our beliefs, but mere causation is not sufficient for justification; 2) even if experience of "brute facts" gives us awareness of "bare particulars," it does not give us the perspective to judge the *authority* of our evidence, the relevance of the observation of a particular to propositional judgments involving the application of concepts. The upshot is that "nothing can count as a reason for holding a belief except another belief" (Davidson 1983, 426). As a result, Humean skepticism about the justification of our inductions is still a looming threat.[64] Clearly, then, if we are to dispel this threat, we must find an answer to the two Sellars-Davidson points.

[64] Coherentists of this stripe are of course quite content with this result, and do not think skepticism is necessary, provided that any resulting circularity is "virtuous," or any infinity is "progressive." This is no place to assess the case for non-skeptical coherentism, of course. We can only observe that the case is a difficult one to make, and that some of its most ambitious advocates in recent years, such as BonJour (1985), have given up the fight (BonJour 2000).

Notice, however, that in the course of searching for solutions to Humean doubts, while we have looked to scientific practice for an account of inductive justification, we have not yet located the source of Humean *doubts* in scientific practice. In fact it is quite plausible that doubts about the epistemic relevance of the empirically given—even if not in the same form as Sellars and Davidson—were also relevant in Humean skepticism about induction. Hume's sensationalism, for instance, entailed the impossibility of perceiving causal relationships. We will soon see that aspects of his view of the senses had scientific sources, but as our exploration deepens we will see that more and more philosophical presuppositions begin to creep in. In our next and final section, I show that our anti-skeptical strategy of making free use of science to answer skeptical reductios can pay some dividends even on the question of skepticism about the senses. There is a substantial body of scientific evidence, particularly in the field of psychology, which can help dispel the doubts summarized in the two Sellars-Davidson points, and others relevant to Hume's views about the perception of causal relationships. But these dividends are limited. Science will only take us so far; we will have to make some normal philosophic arguments to dispel doubts past a certain point. But by the same token, these same doubts are not the product of science. This will suggest that Quinean naturalism's roots in the Humean predicament are not themselves fully naturalistic.

Clearing the naturalistic ground for inductive foundations

In this section, I address each of the two points of the Sellars-Davidson objection to empirical foundationalism, by indicating the kind of scientific evidence they seem to derive from and the wider body of evidence that should cancel their significance. In particular, I show how the wider evidence answering the first objection also sets the stage for addressing the second. In both points, however, we will find philosophical assumptions buried beneath scientific concerns.

I begin with the first anti-foundationalist objection: that experience cannot help justify our beliefs because at best it can only help *explain* them, causally. As with previous skeptical problems, there is an important kernel of truth here. Anti-foundationalists are correct to think that *mere* explanation would probably not suffice for justification. So the question is whether experience *merely* causally explains our beliefs, or perhaps does this plus something more.

A further kernel of truth in the first anti-foundationalist objection is that the conventional empiricist picture of experience does lend credence to the idea that experience merely causes our beliefs. This can be seen in both classical and modern versions of empiricism. In the classical empiricism of Locke and Hume, experience is conceived as an awareness of "ideas" or "perceptions" in an inner mental theater. At best experience offers an indirect, representational awareness of the world if there is some inference one can make from the inner mental theater to the external world. Famously, of course, Berkeley and Hume show that such an inference is difficult if not impossible or incoherent. Modern empiricists (the logical positivists) dispensed with the idea of internal mental entities, and instead sought to define or translate knowledge claims in terms of sensory *language.* This met with failures we have already discussed. In this regard Quine was to the positivists as Hume was to the earlier classical empiricists, drawing as he did skeptical conclusions from shared basic premises of his predecessors.

Both Hume and Quine were informed in this effort by the traditional empiricist view about experience. Quine thought that empiricism advanced with the focus on language rather than "ideas" (1975b), but interestingly, the classical empiricist view was not so far removed from Quine's (1969a) idea that the "stimulus meaning" of an observation statement is determined by "certain patterns of irradiation in assorted frequencies" (1969a, 83), or by "the impact of light rays and molecules on our sensory surfaces" (2004d, 288). The classical empiricists did not arrive at their theory of the inner mental theater through sheer fancy, but in large part from scientific considerations also reflected in Quine. Yolton (1979) presents the widely accepted view that classical empiricists' theory of experience was in large part influenced by their reading of early modern theories of optics and vision. In the *Treatise*, for example, Hume states that "my senses convey to me only the impressions of colour'd points, dispos'd in a certain manner" (1978, 34). This and other passages (in Hume and Berkeley) bear close resemblance to passages from Newton's *Opticks*.[65] Elsewhere Hume makes reference to optical analogies popular in his day, such as those involving "objects seen in mirrors, the eye compared to a lens, the understanding compared to a *camera obscura*," etc. (Yolton 1979, 210). Hume even agrees with the claim of Berkeley's theory of vision that our sight could not give us information about such things as the distance or extension of objects (Hume 1978, 56, 191).

Of course some of this should not be surprising to Quine, who makes reference at least to the issue of the scientific sources of skepticism about three-dimensional awareness (1974, 1–2). But he does not seem to appreciate just how deep the roots of this skepticism run. In response to the contention that we could not be aware of three dimensions if we begin with a two-dimensional retinal image, Quine responds that it is of course by examining three-dimensional eyes that we first discover the two-dimensional retinal image. This response shores up his naturalism, but it concedes that the retinal image—or more generally, mere sensory *stimulation*—is our only source of information about the world. It is this "proximal stimulus" view that Davidson (1990a) identifies as one of the chief sources of Quine's own skepticism. Davidson suggests that a "distal stimulus" view, one which locates the source of meaning and evidence in the "very events and objects that [observation sentences] are naturally and correctly interpreted as being about" (72) would not have the same skeptical tendencies.[66] In spite of this, the "proximal stimulus" view is intensely popular among empiricists, and may even sound uncontroversial at first blush to the modern reader.

Nevertheless, there are serious scientific (and philosophical) grounds for doubting the "proximal stimulus" view, which holds the retinal image as the paradigm of perceptual information. A persuasive defense of a "distal" theory of perception was famously advanced by psychologist J.J.

---

[65] Newton speaks of rays of light conveyed "in so many Points in the bottom of the Eye, and there to paint the Picture of the Object upon that skin" (Newton 1952, 15)

[66] Davidson, of course, does not see any salvation for foundationalism, even in the "distal stimulus" view. He only sees a source of meaning and evidence, not justification. Justification, for him, must come through coherence. In what remains, I will show how a decent "distal stimulus" theory will actually provide grounds for foundationalism not only about meaning and evidence, but justification as well.

Gibson in the mid-to-late twentieth century (1966; 1986). Gibson insisted that a theory of vision is not exhausted by a theory of optics, and that theories treating isolated stimuli as the only source of perceptual information are impoverished accounts of the biological nature of perception. In his theory of vision, Gibson replaces the retinal image with his concept of the "ambient optic array," the totality of structured light surrounding the organism, which, he argues, uniquely specifies (i.e., isomorphically covaries with) the layout of the organism's environment. Perception of objects in an environment works through "information pickup" from this ambient array, via the organism's *active* interaction—involving all of its sensory modalities, in the context of its total bodily motion—with that light. Through this active interplay, the organism discovers the "invariant" properties of the world, for example the invariant relationship among the different angles of a solid object. The organism is not aware of the light, but through a physiological process, *uses* information in the light to achieve (not infer) awareness of its environment. Gibson originally formulated his theory in the course of his research on the visual perception of WWII fighter pilots, whose distance-perception suffered from anomalies which could not be understood using the classical empiricist ideas of "cues" for depth perception. But he gradually assembled an impressive array of further experimental evidence (1986, 147–202). His theories spawned an entire discipline, ecological psychology, which continues active research to this day. This research is now adding steam to work in the philosophy of perception, where more and more philosophers are beginning to take up the cause of "direct realism" (Pollock 1986; Pollack and Oved 2005; Kelley 1986; Putnam 1994; Huemer 2001; Noë 2002). According to direct realism, perception of the world is "direct" in that involves neither awareness of mental intermediaries nor inferences drawn from them. Of course Gibson and direct realism are not without their critics, but the strength of the research program and increasing popularity of its arguments suggest that the epistemic significance of direct perception is not to be brushed aside.

Direct realism is, of course, a controversial philosophical position. But it is a position which did not exist in its current state of development at the time Quine pondered the Humean predicament and its seemingly unavoidable anti-foundationalist consequences. Many traditional objections to direct realism have since been answered (see Le Morvan (2004)), and I do not wish to restate these here. What is important, I think, is that this *is* a philosophical debate, even if there are also scientific issues it brushes up against. In the next few pages, I will mention some of the scientific and philosophical issues that arise in the debate, with emphasis on the latter. I will not be able to answer every objection to direct realism, and that is not the point. The point is that there is a philosophical (though scientifically informed) theory available that helps to answer the first of the Sellars-Davidson points. Perception surely is a partial cause of our beliefs, but the Gibsonian point is that it is not *merely* a cause. Justification may be a species of causation, provided that the right kind of causation is involved. If Gibson and others are right, perception is a form of cognitive awareness of the world. Insofar as the world is what our beliefs are about, at least part of the ground is cleared for seeing how perception might also help justify beliefs about the world, and how the second of the Sellars-Davidson points might be answered.

Probably the most influential critique of Gibson, on both scientific and philosophical grounds, was advanced by Fodor and Pylyshyn (1981). They claim that even though there are important elements of truth in Gibson's critique of predecessor theories of vision (such as Helmholtz's), his "direct perception" cannot be as direct (non-inferential) as he would like. They argue that the subject must make use of *inference* from the ambient light array in order to determine properties of the environment. It will be useful to briefly review Fodor and Pylyshyn's objections, not only to see how Gibsonians might answer them, but also to see how they depend on philosophical, rather than scientific presuppositions.

Fodor and Pylyshyn argue that Gibson's view of information pickup trivializes perception, unless can find a way to constrain information pickup to the specific properties of the environment he mentions, such as "texture, shape, illumination, reflectance, and resistance to deformation" (144). Gibson characterizes these as "ecological properties," but Fodor and Pylynshyn argue that Gibson has no non-circular way of defining which properties count as "ecological." In particular, they allege that he cannot define the ecological properties, because there are no "ecological laws" in terms of which these properties might be understood as being "projectible." They block another route to understanding property detection in terms of the individuation of perceptual systems which respond to them, arguing that Gibson's attempt to individuate these systems in terms more holistic than individual modalities is lacking an independent criterion.

In fact I think that Gibson could indeed try to define ecological properties in terms of ecological laws, but he would in turn define ecological laws in terms of the ultimate functions of perceptual systems understood in the most holistic sense. He would say there is only *one* perceptual system: the perceptual system of the living, acting, organism, and its function is to aid the organism in its survival. Ecological properties, then, are those aspects of the environment which bear most directly on the survival of the organism. This fairly straightforward biological significance is something that Fodor and Pylyshyn's computationally-oriented theory, like so many others, seems to ignore. Turvey et al. (1981) elaborate upon this error, showing how the "affordances" of objects—their dispositions to be grabbed, climbed, dug into, copulate with, crawled upon, lept over, alighted upon, etc.—are the properties which information pickup responds to, *because* in the evolutionary history of the animal, these are the properties that have been most crucial in the organisms success at grabbing, climbing, digging, copulating, crawling, leaping, alighting, etc (1980, 260–7).

Fodor and Pylyshyn's remaining objections are more philosophical than they are scientific. To begin with, they argue that Gibson cannot account for what *they* take to be paradigm cases of perception, such as perceiving that something is a shoe or that something is a Da Vinci. But these cases are paradigmatic only on the "Establishment" view that treats perception as a form of judgment or "perceptual belief," according to which any justifiable "perceives that" locution counts as a *perception* in the strictest sense. According to Gibson, however, what makes perception "effective" is not necessarily that it results in a belief, but more fundamentally that it involves and guides activity. Gibson is not committed to the idea that perceptual awareness is in propositional form.

The view that perception comes in propositional form is, of course, endemic in philosophy, and Fodor and Pylyshyn support it by insisting that if perception were not propositional, we would have the "problem" of being unable to account for "misperception," i.e. sensory illusions and deception. To the naïve reader, avoiding the possibility of sensory deception would seem to be a strength, not a weakness, of a theory of perception that seeks to account for the possibility of foundational justification. One way of responding to the argument from illusion is that while we do often commit errors in *judgment* regarding the stick in water, this does not mean that perception itself misrepresents the object. Propositional content is not needed to account for what the illusion of a bent stick has in common with an actual bent stick. All that is needed is to say that each of these involves a similar perceptual look (Turvey et al. 1981, 275).[67] An illusory percept has this look in common with "veridical" percepts of actually bent sticks, but is illusory only in that it is an unusual form of perceiving a stick that is actually straight.

There is, of course, a philosophical dispute about whether illusions can even be described as illusions without implying that they involve misrepresentation. The usual locution is that an illusion involves seeing a straight stick *as* a bent stick, and since the straight stick is not bent, this implies a representation of something as something that it is not. Of course it is important to distinguish two senses of "seeing a straight stick as a bent stick." One sense involves applying the concept "straight" to the stick, which is the propositional form of awareness we are not at the moment considering. The

[67] It is true that certain forms of perceptual awareness will deliver more or less information, which may, in a given situation, have more or less biological utility. But this does not mean that the senses are "mistaken." First of all, biological disvalue is not the same as epistemic disvalue. Simply because some forms of awareness are out of the ordinary does not mean they fail to be forms of awareness of the world. It is possible to grasp the same objects in different forms. Second, what appears to be a biological disadvantage in a narrow circumstance can be understood as a consequence of the possession of a biological capacity that is itself of great utility. The ability to detect refraction differentials, for example, while misleading to a subject's judgment in certain cases, is nonetheless essential to the organism's ability to detect the difference between solid, liquid, and air—a crucial ability from the perspective of survival.

remaining sense, then, is simply a comparison of the two perceptual appearances. How we are to understand perceptual appearance, of course, is a difficult philosophical question. If appearances are *qualia*, which many philosophers have difficulty reducing to the physical, it may be difficult for external observers even to compare appearances to objects in order to say that the two are different and that *qualia* misrepresent. And as Berkeley argued long ago, it is impossible for the observer himself to get outside of his consciousness and compare his ideas with things outside his ideas. (Of course we can agree with Berkeley about this, without following him to his idealist conclusions, if we remember that *qualia* are our forms of awareness, not the *objects* of awareness.) Now of course there may be physical correlates on which *qualia* supervene, and perhaps these may be compared to the objects themselves. The big question, however, is whether there is a naturalistic concept of representation that could sustain a fruitful comparison. I have discussed problems with this project in some detail in chapter 3. The most viable theory is that representations are somehow isomorphic with what they represent in the world. One problem with applying this view to the question of illusion is that arguably an illusory representation of a straight stick as bent is still isomorphic with the straight stick on some way, because isomorphism is cheap. The causal connection between the two virtually guarantees this. Surely the illusory representation does not have the same isomorphism with a bent stick as a non-illusory representation, but then the question becomes, why privilege one isomorphism over the other?[68]

A final objection to Gibson from Fodor and Pylyshyn is that even if Gibson can specify the detected ecological properties that bring about perception, it is only through a process of *inference* that one can proceed from these detected properties to perceiving the environment. If perception requires inference, it is clearly not direct, and for that matter, not clearly distinguishable from judgment, and

[68] Turvey et al. (1981) make a useful response to the objection that direct realism is somehow incompatible with the possibility of illusion. They point out that to say that the senses are "in error" suggests that they should represent the bent stick in water (for example) differently than they actually do. But *this* ignores the fact that the senses' response to the structure of light in a way that displays the "differential in refractive indices between the media of air and water" actually delivers important information to the organism, and that there is no standard by which to judge this report as "in error."

therefore subject to error. This inferentialist view of perception is not unique to Fodor and Pylyshyn's particular brand of computationalism, but extends back as far as the perceptual theories of Helmholtz.

It is undeniably true that a great deal of physiological processing intervenes between the reception of stimuli (globally understood) and the production of percepts. The question is whether there is good reason to characterize this physiological process as a process of *inference*. This question is parallel to the question of whether or not to consider physical correlates of *qualia* to be representations. Arguably a naturalistic account of perceptual inference depends on a naturalistic account of representations, for surely naturalists would admit that inference is not inference unless it involves the manipulation of representations. So all of the problems raised above and in chapter 3 count against this position, as well.

There are a number of other problems with characterizing mere physical processing as "inference." Our usual conception of inference would seem to commit the inferentialist to the idea that pre-conceptual inference would require innate knowledge, and also the possibility of reversing illusions. I will not elaborate on these at present.[69] There is one additional objection from the commonsense concept of inference that is worth focusing on, for the sake of bringing out some of the purely philosophic points at issue here. An inference is usually understood as a kind of argument drawn from known premises. The "premises" of inferential perception would, presumably, be either

[69] Kelley (1986) points out two further problems with the inferentialist view of perception. First, inference is also usually understood as involving the application of background knowledge to known premises. However the kind of background knowledge presupposed by the sorts of inferences allegedly needed for the transition from detected properties to perception would be of an extremely advanced kind. In order to calculate three-dimensional perception based on ocular disparity, for example, the pre-conceptual child would need knowledge of the geometry of parallax. Since newborn children can perceive three-dimensional objects—but have not yet studied geometry—this knowledge would need to be innate. Even if this kind of innate knowledge is not impossible, the burden of explaining its possibility is nonetheless a heavy one. Direct realists, by contrast, can accept any physiological findings relevant to the mechanism of perception, while simply insisting that whatever the mechanism is, it is not inferential. See Kelley (1986, 61). Second, inference is usually understood as a conscious mental process. The kind of inference described so far—involving premises and background knowledge of which the perceiver is not aware—would have to be a special kind of unconscious inference. The discovery that the conclusion of a conscious inference is false often leads us to reassess its premises and reject the same inference in the future. Yet if perception is inferential, this means that perceptual illusions are as well. Only we cannot "reverse" perceptual illusions as we might reverse other "erroneous" inferences. We do not stop seeing the stick as bent when we learn that it really isn't. See Kelley (1986, 67).

the properties of the retinal stimuli (in the proximal viewpoint) or the detected ecological properties (in the distal viewpoint). Surely we are aware of objects *by means* of retinal stimuli or surface properties, but the question is whether this implies *awareness* of these properties *as* properties. The proper direct realist response would be to suggest that the means of awareness and object of awareness should not be confused, so awareness of objects by means of retinal stimuli and surface properties does not imply awareness *of* these stimuli or properties, and therefore does not imply the use of that knowledge as premises in an inference. Now there is a long tradition, most pronounced in the Kantian tradition, which holds that only *absolutely* unmediated awareness counts as awareness. This is what prompts philosophers to look for the objects of awareness in the most "internal" states possible, rather than in external objects. But there is simply no good philosophical reason to think that awareness counts as awareness only if it is unmediated. Hence there is every reason to take phenomenological facts at face value: we are aware of objects, not images or stimuli or even surface properties.[70]

Even if Gibsonian direct realism is plausible, another part of that ground remains to be cleared. As BonJour has noted in his recent critique of direct realism (2004), even if we accept that perception is a form of awareness unmediated by inference from internal mental objects, this does not yet show how it is related to the justification of beliefs. Particularly if we are to understand direct perception as a non-conceptual, *non-propositional* form of awareness, there is then a serious problem about how such a form of awareness could come to justify conceptual, propositional beliefs.[71] BonJour considers Reynolds' (1991) version of this view, according to which non-propositional experience justifies by being the object of recognitional skills that can be captured in rules but that are not articulated as such by the recognizers. BonJour is concerned, however, that Reynolds' view of

[70] See Kelley (1986, 66-8).

[71] Considering this version of direct realism is clearly advantageous, as it avoids the objections BonJour raises to the other versions which consider perception to be either propositional but non-conceptual Huemer (2001) or both propositional and conceptual Brewer (1999). The objection to both such views is, essentially, that if perception is propositional, it could be true or false, and is then in need of justification itself. This means it cannot serve the role of a foundation that is usually demanded of it. Of course I think the answer here is that perception is non-propositional, and the automatic deliverances of the senses are neither true nor false. See paragraph 4 of footnote 16 for more.

justification is not sufficiently internalist: according to it, we follow these recognition rules only because our "epistemic community" has adopted them for their efficacy in producing true beliefs—it is not clear if we adopt them because of our own understanding of their connection to truth (Bonjour 2004, 365). As before, this skepticism about the efficacy perception stems from a core of truth: there does indeed seem to be a difference between merely learning to replicate a social skill and forming a justified belief. If we are to defend the epistemic significance of direct realism, more is needed.

One possibility is that we can understand perception's epistemic relevance to propositional justification via *concepts*. Justified judgment essentially means the justified application of concepts to a new situation. Of course there is a longstanding philosophic tradition, deriving from Kantian-Hegelian premises, that takes a dim view of subject-predicate judgment as a source of genuine cognition. This view finds its modern equivalent in the doctrines of Sellars and Brandom, who allege that an organism's ability to categorize its environment makes it sentient, but not sapient. To cognize is to do more than to have a reliable differential response to the environment. Cognition is instead a product of the inferential commitments of one's judgments, commitments entered into through social interaction. It is of course now obvious that we are dealing with a *paradigmatic* non-naturalistic philosophic objection to the possibility of foundationalism. If naturalized epistemologists have to rely on "myth of the given" objections to counter foundationalism, they are in a great deal of trouble, because these skeptical doubts are not prompted by any obviously scientific problems—unless it makes sense to say that Kant's and Hegel's theories were so prompted.

Nevertheless I want to say a few things about how to answer BonJour's concern in a way that addresses the problems raised by Sellars and Brandom. Of course one issue is the worry about the difference between sentience and sapience. I hope that I have already answered that objection to a degree by showing that perception is not a mere response to the environment, but an active, cognitive probing of it. This doesn't do everything to explain the possibility of *human* sapience, of course, because even animals have perceptual capacities. A question remains as to how perception gives any authority to subject-predicate judgments.

John McDowell, a partisan for the Sellars-Brandom critique of the authority of the empirically given, considers one way in which perception might be taken as authoritative. We may be able to understand the conditions of justified concept-application in relation to the conditions in which the same concepts were originally formed: we are justified in applying a concept whenever we observe aspects of our environment to be similar to the aspects of the environment from which we *formed* the concept. McDowell considers just this possibility in order to dismiss it:

> We could not begin to suppose that we understand how pointing to a bit of the Given could justify the use of a concept in judgment—unless we took this possibility of warrant to be constitutive of the concept's being what it is, and hence constitutive of its contribution to any thinkable content it figures in. . . . The supposed requirement is reflected in a familiar picture of the formation of such concepts, a picture that is a natural counterpart to the idea of the Given . . . [that] the associated conceptual capacities must be acquired from confrontations with suitable bits of the Given. . . . A subject would have to abstract out the right element in the presented multiplicity (1996, 6–7).

So McDowell acknowledges that an abstractionist view of concepts could help explain how "the Given" justifies judgments. But invoking the work of Geach (1957), he simply denies that observational concepts *could* be abstracted. Now Geach offers compelling reasons for why *color* concepts could not be the first ones we abstract from perceptual experience, contrary to traditional empiricists. Indeed he is right that young children form concepts of middle-sized objects well before they form color concepts (or any other attribute concepts, for that matter) (Anglin 1977, 11). But at no point do Geach or McDowell offer reason to doubt that *these* concepts of middle-sized objects might be formed first by abstracting from perceived similarities. It is worth considering if that is possible.

John Pollock and Iris Oved (2005) understand that we recognize middle-sized objects, like cats, even when we cannot articulate a description of cats in terms of shape- or color-concepts. They speculate about how this recognition might result from a "cat detector" cognitive "module," that is acquired after seeing how the various parts of cats fit and move together in a certain uniform way, in just the same way that chicken-sexers learn their skill on the basis of an implicit learning of chicken-parts. Perhaps this results from the formation of a connectionist "neural network" that permits a sophisticated form of pattern recognition (333–8). What Pollock and Oved are gesturing toward is the recognition that there is a *perceptual similarity* relation available for three-dimensional shapes. Even though young children do not yet understand shape concepts—especially not those of the three-dimensional kind—it does seem that the most salient similarities among the first kinds of objects they do conceptualize (types of animal, people, food, artifacts, etc.) are in regard to shape (Anglin 1977, 71). It is important that this similarity is perceptual—meaning non-propositional and non-inferential—for otherwise we fall prey to Sellars' objection that a *judgment* of similarity would presuppose further concepts (such as that of "similarity," and that of the respect in which things are said to be similar) (Sellars 1963, 160).[72]

There is, however, still something of a puzzle about how such three-dimensional similarities could be grasped without the possession of any prior concepts. This is where Gibson's view of perception begins to pay dividends even for our understanding of concept-formation, where our answer to the first Sellars-Davidson point informs our answer to the second. Recall that central to Gibson's theory is the distinction between information that is "picked up" from the ambient optic array on the one hand, and the things or properties of the world that are perceived on the other. Information about properties such as shape and resistance to deformation are picked up, and "invariant" relationships among these are extracted in a way that produces an awareness of the object as a whole. So something like the grasp of three-dimensional shape is already involved in perception of a given object in the first place: it comes as little surprise, then, that this ability can be marshaled for the purpose of grasping similarity relationships.

[72] These similarities would need to grasped relationally, by seeing how two objects differ less from each other than they do from a third. They could not result from simply reading off common features from any two objects, in the Lockean fashion, for this would presuppose a conceptualization of the features in question. For more, see Kelley (1984, 336-342).

Jesse Prinz (2002) has formulated a new theory of "concept empiricism" that explains how concepts of middle-sized objects are abstracted from perceptual experience, which exploits mental devices he calls "proxytypes," products of "long-term memory networks" which permit the grouping together of quantitatively-varied percepts. Simple examples of these networks include "hierarchical" representations, which encode and integrate changes in experience as one "zooms" in or out on an object, "transformational" representations, which do the same for moving objects, and most importantly, "*predicative"* representations, which exploit the other types to group objects on the basis of similarity (as if a similarity range could be grasped as a transformation from one similar to another) (141–4).[73] Here, in particular, we see how proxytype theory predicts just what we would expect about similarity in light of Gibson's findings about perception:

> Proxytype theory offers an attractive account of basic-level categorization. The basic level generally lies at an intermediate [middle-sized] level of abstraction. . . . This suggests that shape similarity confers an advantage in learning and categorization. Proxytype theory predicts this because proxytypes are perceptually derived and shape plays a very dominant role in object perception. (163) [74]

Now Prinz's theory is somewhat controversial, but it does have the advantage of cohering with a vast amount of evidence from developmental and cognitive psychologists about concept-formation and use, which previous theories seemed to be able to deal with only piecemeal. His theory is also incomplete: it lacks, for example, a better account of how awareness of similarity makes possible a

---

[73] See also Kelley (1984), 345.

[74] It is also worth noting how similarity understood in light of Gibson and Prinz helps answer long-standing objections to similarity as the basis for an awareness of natural kinds, as expressed by Quine (1969b). Quine says there is something "logically repugnant" about similarity, understood in the comparative fashion mentioned in footnote 18, because of the inability to identify the properties in terms of which this similarity should be formulated. Quine repeats the familiar point that two objects are similar in any number of regards, and similarity threatens a promiscuity of natural kinds unless the relevant respects of similarity are constrained. Quine speculates that standards of similarity may be innate and determined by evolution. Here we can agree with him, but the types of similarity standards he assumes to be unique are primarily color-similarities, which is an age-old empiricist bias that has long been contradicted by developmental evidence. Children are far more impressed by similarities in shape than they are by color, and we can now understand *this* in evolutionary terms by reference to Gibson's understanding of ecological properties as forming the basis for the information pickup of perception. And when similarity is understood first in terms of shape, similarity also becomes much more respectable, scientifically. The shapes of objects have far more consequences for their causal interactions, which makes concepts formed on the basis of perceptual similarity "natural kind" concepts from the beginning, not just superficially-held "nominal essences." Science itself, then, to the extent that it begins, developmentally, with similarity concepts, is not, as Quine says "rotten to the core."

unitary grasp of seen and unseen referents of a concept. In fact I critiqued Prinz in chapter 3 for offering his theory as a purely naturalistic account of concepts. Perhaps the most overwhelming reason to consider mine to be a non-naturalistic approach is because I do not attempt to offer a naturalistic solution to the "*qua*" problem: the problem of which of many classes of things a categorized object is represented as belonging to. I raised that problem for higher-level concepts which proxytype theory could not address by itself, but it is worth mentioning that even proxytypes face their own version of the problem. Even though the mind *can* transform images in long-term memory networks in the ways described above, there is a question about whether it ever *does* transform them into all of the possible objects we take our concepts to refer to. Perhaps a naturalistic theory could be formulated in terms of dispositions to transform proxytypes, but disposition talk is often modally loaded and difficult for the naturalist to reconciling. It is here that I think that the first elements of irreducible intentionality may enter the picture. Perhaps part of what permits us to refer to an infinite range of objects is little more than our intention to regard the range of transformations as possible, and our intention to classify potential objects under that range.[75] This irreducible intentionality would also help erase the Sellars-Brandom concern that subject-predicate judgment involves mere differential response to one's environment. Since I am presenting a non-naturalistic theory of reference, it should be clear that we are not talking about merely differential response.

We are now approaching our dénouement. We have seen how scientific evidence, and some philosophic theorizing, lends credence to a theory of perceptual similarity that helps explain the abstraction of basic observational concepts from perception. We needed such a view of abstraction as an aid for understanding how perception could help justify judgments, in the manner McDowell thought to be impossible. We needed such a clarification about judgments to answer BonJour's concern that, even if perception did afford direct awareness of the world, it would not be such as to justify our judgments. To answer BonJour's concern, and illustrate how all concepts used in judgment

---

[75] See Rand (1990) for a developed version of this account.

might be justifiably applied, we need an account of concepts beyond simply the middle-sized object-concepts discussed so far. Particularly if we want an account of judgment that allows for inductive generalization, we need some account of action and attribute-concepts, one showing how these might be derived from perception.

In Gibsonian theory, not only are middle-sized objects or entities perceived, but some particular attributes, actions, and relationships are as well. (Other recent psychological work influenced by Michotte (1963) even suggests the possibility of perceiving *causal* relationships.[76] The perceiver perceives a total environment, and to the extent that any types of entity, attribute, action, or relationship have played an ecologically significant role in the organism's evolutionary history, it makes sense that the organism should come to develop a capacity to perceive them, i.e. to be able to *discriminate* one particular entity from another entity, one particular attribute from another, etc. To perceive an object is just to be able to see it as distinct from other objects against a shared spatial background. Perception of other properties is also relational in this way (contrary to classical empiricists): one does not perceive the particular shape of an object, for example, by discriminating its shape from its color, but instead by discriminating its shape from another object's shape. Perceptual similarity can be understood in terms of perceptual discrimination: two objects, for example, are perceived as similar if the respect in which they are perceived as differing (three-dimensional shape) is less than the respect in which they each differ from some third outlier. By a parallel act of double discrimination, we could also come to perceive similarities in attribute, action, or relationship. We can perceive similarities in shape if, for example, we perceive two objects of different types each of which is spherical (e.g., an orange and a baseball), as against everything else. Or we can even perceive similarities in action if, for example, we perceive two objects of different types each of which is rolling (again the orange and the baseball), as against other objects at rest.

[76] See Leslie and Keeble (1987), Scholl and Tremoulet (2000), and Prinz (2002, 173-177).

Once we appreciate the way perceptual similarity can be extended even to categories apart from entities, forming concepts from these similarities could proceed using at minimum something like a Prinzean proxytype. Applying these concepts in the act of judgment would then yield generalizations ("balls are round," "balls roll," etc.), seemingly capable of justification by reference to perceiving the same similarities that first permitted the formation of these concepts.[77]

We were first led down the path of looking for the foundations of induction by John Norton, who argued that as long as one inductive generalization could be understood as justified by a material fact induced by another, there need be no infinite regress of inductive justification as long as we appealed to more and more basic facts, eventually terminating in "brute" facts of experience. We have now gone some distance in explaining just what these "brute facts" might look like (literally). But we are now also in a position to appreciate and enrich a tantalizing suggestion of Norton's (2003, 668) in a final footnote:

> I reject the simple argument that such brute facts are always singular and that no collection of singular facts can license a universal. The problem is that canonical singular facts—"the ball is red"—already presuppose universal knowledge. In this case, it resides in the recognition that the thing is a ball and that its surface is red, thereby admitting recognition of commonality with potentially infinite classes of objects and colors.

We can now see just how true this is. The very formation of concepts is, in a way, a quasi-inductive process, inasmuch as we must abstract from similarities to refer to an open-ended class. The application of concepts in judgment therefore presupposes this universality.[78] This also enriches a suggestion from Lange (2004, 218):

[77] On non-Humean theories of causality, being able to generalize about the actions of entities (e.g., "balls roll," "knives cut") already constitutes causal knowledge (Harré and Madden 1975). On this view, causal connections are not primarily regularity relations between types of event, but relationships between types of actions and types of entities. It is worth noting that an additional source of Humean problems is the view that causal relations are only among events—where the events are interpreted psychologically as sensory atoms. With the rejection of the retinal image view of visual perception, however, the Humean metaphysics of events loses its psychological grounding. See Harré and Madden (1975, 49-67).

[78] It may be objected—following Hume (1975, 33)—that even if we possess general knowledge connecting subject and causal predicate, we do not know when to apply it to particulars, and hence we can never make predictions about singular concretes in the future or in unobserved distant locations. Hume's idea is that there are no known connections between "sensible qualities" and "secret powers," and so merely recognizing objects

> If we cannot make observations without undertaking commitments that go beyond them, then apparently the inductive skeptic must abstain from making *any observations at all*. Thus the problem of induction is circumvented, since it presupposes that we could be entitled to make observations without having any reason to infer inductively beyond them.

If this is right, we have succeeded not only in stopping the inductive skeptic's *reductio*, but in the course of doing so, we have shown that quite possibly the Humean skeptic about induction does not even have the logical right to assume the premises he did in the first place. He cannot assume the premise that we can formulate certain observation reports about singular facts, which are assumed not to imply generalizations, as there would be an outright contradiction involved in the idea that we might make observation reports but not imply any generalizations. This gives reason to wonder whether many of the other premises assumed for reductio might not involve similar contradictions with the skeptic's conclusion. It also gives further reason to wonder about how many of these doubts are prompted by purely scientific discoveries, rather than by philosophic presuppositions.

---

similar to acting objects in the past does not guarantee that they will act the same way or possess the same powers. There is some truth to this point: grasping a similarity is certainly only a defeasible source of justification, and we can make mistakes. But at this point, Hume is at the end of his rope. The objection that we might be mistaken about the application of a generalization to a particular is no longer a special problem for induction. It is really no different than the simple argument from illusion concerning the bent stick in water. Just as we might be subject to shape illusions, so we might also be subject to causal illusions. But this is no special problem for our knowledge, provided that we understand that the error is one of judgment, not the senses, and that it is the senses which provide the ultimate source of understanding and correcting our errors. Just as we can remove the stick from water to see if it is really bent, we can place our bread under a microscope to discover if it has the ordinary chemical sub-structure that would nourish us. Of course we are subject to error when we make judgments about these observations, too, but the general possibility of error does not imply that we might be wrong in any particular case. ("It is possible for me to err" does not imply "It is possible that I am in error now.") Here we are really treading on the grounds of Cartesian skepticism, rather than Humean skepticism. The most I can say is that the manner in which the Cartesian skeptic generates possible sources of error is akin in many important ways to Quine's underdetermination problem. Just as multiple rival theories "possibly" might account for the same observational evidence according to Quine, there might also be different "possible" possible sources of error regarding evil demons, etc., might account for what we perceive. I think each of these arguments is flawed for many of the reasons I have already identified in my section on the underdetermination thesis.

Conclusion

This essay began by considering a strategy, sometimes attributed to Quine, for refuting skepticism by pointing to the scientific assumptions made by skeptics for the sake of *reductio*: If the skeptic has the right to assume science, then the anti-skeptic also has the right to assume science to defuse the absurd consequences of the skeptic's more scientifically isolated assumptions. This strategy does not itself attempt to prove any propositions of philosophy, but simply to dismiss motivations for various kinds of skepticism about philosophical propositions. I have sought to show how this strategy may be trained upon more an even radical skepticism than has been imagined, including Quine's own form of skepticism, the underdetermination thesis. But the underdetermination thesis was only a default position accepted as a result of previous skepticisms about induction and empirical foundations, so I sought to show how a similar anti-skeptical strategy can be applied against these positions, as well.

I have argued above that the Quinean anti-skeptical strategy of appealing to more science does have some limited utility in dissolving Humean doubts about induction, and Sellars-Davidson doubts about the perceptual basis of induction. But my emphasis, particularly in this last section, has been that scientific evidence cannot do the whole job, because not all of the doubts are themselves scientific. Even if we cannot resolve these doubts, my concern then is that naturalized epistemology cannot rely on these doubts to motivate its program. If skepticism seems unavoidable only because philosophers have idiosyncratic presuppositions about mind and representation, then insofar as skepticism prompts naturalism, naturalism is prompted for non-naturalistic reasons. This works to undermine the coherence of the naturalistic view that philosophy works in a manner that is purely continuous with the scientific. If purely philosophical problems got us into skeptical trouble, it may take philosophical solutions to get us out.

But there is some legitimate concern over what a "purely philosophic" solution would be. In the course of proposing some of these solutions, I have made appeals to our understanding of concepts like “perception,” “inference,” “concept” and “intentionality.” I have argued that skeptical arguments fail because of their misapplication of these concepts. Nevertheless I want to emphasize that this methodology needn’t be *a priorist*, even if it is “purely philosophical.” Pure philosophy is only *a priori* if what separates philosophy from science is that only the second involves the empirical method. Other methods of demarcation are possible. It is possible for philosophic knowledge to be empirical, if only more fundamental and universal than scientific empirical knowledge. Philosophy may examine facts—and the concepts that designate them—that are of fundamental important to scientific and all other kinds of knowledge. If we gain better understanding of our concepts of these facts by understanding how they arise from our experience (but extrospective and introspective), rather than through “intuition,” all the better.

Understanding philosophic method this way, while empirical, is also not “naturalistic.” Naturalism takes scientific empirical knowledge as the *exclusive* form of knowledge. My approach certainly takes science as the highest form of empirical knowledge, but not as the exclusive form. Central to the doctrine of empirical foundationalism I have explored, after all, is that there might be pre-scientific commonsense empirical knowledge, on which science itself is based. Philosophic knowledge itself may be of this pedigree, though of a higher order than most pre-scientific knowledge, and certainly more self-conscious. A full articulation of this approach to philosophic knowledge, in particular its empirical approach to “conceptual analysis” is far beyond the scope of this dissertation. Its development will have to wait for another day. But acknowledge its possibility is an important antidote to the view, so often endorsed by naturalists, that the choice is between *a priori* conceptual analysis and scientistic naturalism. This is a false choice, and I hope that much of the work of this chapter—and dissertation—will help to show why.

REFERENCES

Almeder, Robert. 1990. On naturalizing epistemology. *American Philosophical Quarterly* 27 (4): 263–79.

Anglin, Jeremy. 1977. *Word, object, and conceptual development.* New York: W.W. Norton and Company.

Ariew, Roger. 1984. The Duhem thesis. *British Journal for the Philosophy of Science* 35: 313–25.

Aristotle. 1941. *De anima*. In *The basic works of Aristotle*, ed. Richard McKeon. New York: Random House.

Armstrong, D.M. 1973. *Belief, truth and knowledge*. Cambridge: Cambridge University Press.

Bangu, Sorin. 2006. Underdetermination and the argument from indirect confirmation. *Ratio* 19 (3): 269–77.

Bartsch, Karen, and Henry Wellman. 1995. *Children talk about the mind*. New York: Oxford University Press.

Bedau, Marc. 1991. Can biological teleology be naturalized? *Journal of Philosophy* 88: 647–55.

Ben-Menahem, Yemima. 2005. Black, white and gray: Quine on convention. *Synthese* 146: 245–82.

Bergstrom, Lars. 1990. Quine on underdetermination. In *Perspectives on Quine*, ed. Robert Barrett and Roger Gibson, 38–54. Oxford: Basil Blackwell.

———. 1993. Quine, underdetermination and skepticism. *The Journal of Philosophy* 90 (7): 331–58.

———. 2004. Underdetermination of physical theory. In *The Cambridge companion to Quine*, ed. Roger Gibson, 91–114. New York: Cambridge University Press.

Bernier, Paul. 2002. From simulation to theory. In *Simulation and knowledge of action*, ed. Jerome Dokic and Joelle Proust, 33–48. Amsterdam/Philadelphia: John Benjamins Publishing Co.

Binswanger, Harry. 1990. *The biological basis of teleological concepts.* Los Angeles: ARI Press.

Bishop, Michael, and Stephen Stich. 1998. The flight to reference, or how not to make progress in the philosophy of science. *Philosophy of Science* 65: 33–49.

Block, Ned, and Robert Stalnaker. 1999. Conceptual analysis, dualism, and the explanatory gap. *The Philosophical Review* 108 (1): 1–46.

Bloom, Lois. 1973. *One word at a time: the use of single word utterances before syntax.* Janua linguarum, series minor, 154. The Hague: Mouton.

———. 1993. *The transition from infancy to language*. Cambridge: Cambridge University Press.

Bloom, Paul, and Tim German. 2000. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77: B25–B31.

Boghossian, Paul. 1990. The status of content. *The Philosophical Review* 99 (2): 157–84.

———. 1996. Analyticity regained. *Nous* 30 (3): 360–91.

BonJour, Laurence. 1985. *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

———. 2000. Toward a defense of empirical foundationalism. In *Resurrecting old- fashioned foundationalism*, ed. M. DePaul, 21–38.. Lanham, MD: Rowman and Littlefield.

———. 2004. In search of direct realism. *Philosophy and Phenomenological Research* 69 (2): 349–67.

Bonevac, Daniel. 2001. Naturalism for the faint of heart. In *Reality and Humean supervience: essays on the philosophy of David Lewis*, eds. G. Preyer and F. Siebelt. Lanham, MD: Rowman and Littlefield Publishers, Inc.

Boyd, Richard. 1991. Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies* 61: 127–48.

Brandom, Robert. 1998. Insights and blindspots of reliabilism, *Monist* 81: 371–92.

———. 2001. Modality, normativity, and intentionality. *Philosophy and Phenomenological Research* 63 (3): 587–609.

Bretherton, Inge, Sandra McNew and Marjorie Beeghly-Smith. 1981. Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In *Infant Social Cognition*, eds. M.E. Lamb and L.R. Sherrod, 333–73. Hillsdale, NJ: Erlbaum.

Brewer, Bill. 1999. *Perception and reason*. Oxford: Oxford University Press.

Burge, Tyler. 1979. Individualism and the mental. In *Midwest Studies in Philosophy* 4, *Studies in Metaphysics*, eds. Peter French, Theodore Euhling Jr., and Howard Wettstein. Minneapolis: University of Minnesota Press.

Burr, Jean, and Barbara Hofer. 2002. Personal epistemology and theory of mind: deciphering young children's beliefs about knowledge and knowing. *New Ideas in Psychology* 20: 199–224.

Chalmers, David. 1996. *The conscious mind*. New York: Oxford University Press.

Chisholm, Roderick. 1977. *Theory of knowledge*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Churchland, Paul. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78: 67–90.

———. 1996. *Matter and consciousness.* Cambridge, MA: MIT Press.

Crane, Tim. 1998. How to define your (mental) terms. *Inquiry* 41: 341–54.

Cummins, Robert. 1989. *Meaning and mental representation*. Cambridge, MA: MIT Press.

———. 1996. *Representations, targets and attitudes*. Cambridge, MA: MIT Press.

Daniel, Steven. 1999. Why even Kim-style psychophysical laws are impossible. *Pacific Philosophical Quarterly* 80: 225–37.

Davidson, Donald. 1963. Actions, reasons and causes. *The Journal of Philosophy* 60(23): 685–700.

———. 1968. On saying that. *Synthese* 19, 139–46.

———. 1983. A coherence theory of truth and knowledge. In *Kant oder Hegel?*, ed. D. Heinrich: 423–38. Stuggart: Klett-Cotta.

———. 1984. *Inquiries into truth and interpretation*. Oxford: Clarendon Press.

———. 1990a. Meaning, truth and evidence. In *Perspectives on Quine,* eds. R. Barrett and R. Gibson: 68–79. Oxford: Blackwell.

———. 1990b. The structure and content of truth. *The Journal of Philosophy* 88 (6): 279–328.

Davies, Martin, and Tony Stone. 2001. Mental simulation, tacit theory, and the threat of "collapse". *Philosophical Topics* 29 (1&2): 127–73.

Decety, Jean. 2002. Neurophysiological evidence for simulation of action. In *Simulation and knowledge of action,* eds. Jerome Dokic and Joelle Proust, 53–72. Amsterdam/Philadelphia: John Benjamins Publishing Co.

Devitt, Michael. 1996. *Coming to our senses*. New York: Cambridge University Press.

Devitt, Michael, and Kim Sterelny. 1999. *Language and reality: an introduction to the philosophy of language*. 2nd ed. Cambridge, MA: MIT Press.

Dretske, Fred. 1986. Misrepresentation. In *Belief: form, content, and function*, ed. R. Bogdan, 17–36. NewYork: Oxford University Press.

Dummett, Michael. 1973. *Frege: philosophy of language* London: Duckworth.

Ebbs, Gary. 1994. Review of *Pursuit of truth*. *The Philosophical Review* 103 (3): 535–41.

———. 2003. Denotation and discovery. In *Socializing metaphysics: the nature of social reality*, ed. F. Schmitt, 247–68. Lanham, MD: Rowman & Littlefield.

Feldman, Richard. 2006. Naturalized epistemology. In *The Stanford encyclopedia of philosophy (Fall 2006 Edition)*, ed. Edward N. Zalta. http://plato.stanford.edu/archives/fall2006/entries/epistemology-naturalized (accessed July 6, 2007).

Fodor, Jerry. 1974. Special sciences (or the disunity of science as a working hypothesis). *Synthese* 28 (2): 97–115.

———. 1987. *Psychosemantics.* Cambridge, MA: MIT Press.

Fodor, Jerry, and Zenon Pylyshyn. 1981. How direct is visual perception?: Some reflections on Gibson's "ecological approach." *Cognition* 9: 139–96.

Geach, Peter. 1957. *Mental acts: their contents and their objects*. London: Routledge and Kegan Paul.

Gibson, James J. 1966. *The senses considered as perceptual systems*. Boston: Houghton-Mifflin Co.

———. 1986. *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Earlbaum Associates, Publishers.

Gibson, Roger. 1986. Translation, physics and facts of the matter. In *The philosophy of W.V. Quine*, eds. L. Hahn and P. Schilpp, 139–54. La Salle, IL.: Open Court.

———. 1988. *Enlightened empiricism: an examination of W.V. Quine's theory of knowledge*. Tampa, FL.: University of South Florida Press.

Giere, Ronald. 1985. Philosophy of science naturalized. *Philosophy of Science* 52: 331–56.

———. 1988. *Explaining science*. Chicago: University of Chicago Press.

———. 1989. Scientific rationality as instrumental rationality. *Studies in the History and Philosophy of Science* 20 (3): 377–84.

Goldman, Alvin. 1979. What is justified belief? In *Justification and knowledge: new studies in epistemology*, ed. George Pappas. Boston: D. Reidel Publishing Company.

———. 1986. *Epistemology and cognition*. Cambridge, MA.: Harvard University Press.

———. 1993. Epistemic folkways and scientific epistemology. *Philosophical Issues* 3: 271–85.

———. 1995. Interpretation psychologized. In *Folk psychology: the theory of mind debate*, eds. M Davies and T. Stone. 74–99. Oxford, UK: Blackwell. Originally published in *Mind and Language* 4, 161–85, 1989.

———. 2006. *Simulating minds: the philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.

Gopnik, Alison and Janet Astington. 1988. Children's understanding of representational change and its relation to understanding of false-belief and appearance-reality distinction. *Child Development* 59: 27–37.

Gopnik, Alison and Peter Graf. 1988. Knowing how you know: young children's ability to identify and remember the sources of their beliefs. *Child Development* 59: 1366–71.

Gordon, Robert. 1995a. Folk psychology as simulation. In *Folk psychology: the theory of mind debate*, eds. M. Davies and T. Stone, 59–73. Oxford: Blackwell. Originally published in *Mind and Language* 1: 158–70, 1986.

————. 1995b The simulation theory: objections and misconceptions. In *Folk psychology: the theory of mind debate*, in eds. M. Davies and T. Stone, 100–22. Oxford: Blackwell. Originally published in *Mind and Language* 7 (1–2): 11–34, 1992.

————. 1995c. Simulation without introspection or inference from me to you. In *Mental simulation: evaluations and applications*, eds. M. Davies and T. Stone, 53–67. Oxford: Blackwell.

————. 1996. "Radical" simulation. In *Theories of theories of mind*, eds. P. Carruthers and P. Smith, 11–21. Cambridge, UK: Cambridge University Press.

————. 2001. Simulation and reason explanation: The radical view. *Philosophical Topics* 29: 175–92. http://www.umsl.edu/~philo/Faculty/Gordon/MindSeminar99/papers/Gordon/simrex.html (accessed July 10, 2007)

Harré, Rom and Edward Madden. 1975. *Causal powers: a theory of natural necessity*. Oxford, UK: Basil Blackwell.

Heal, Jane. 1995. Replication and functionalism. In *Folk psychology: the theory of mind debate*, eds. M. Davies and T. Stone, 45–59. Oxford,UK: Basil Blackwell. Originally published in *Language, mind and logic,* ed. J. Butterfield. Cambridge: Cambridge University Press, 1986.

Henderson, David. 1994. Epistemic competence and contextualist epistemology: Why contextualism is not just the poor person's coherentism. *The Journal of Philosophy* 91: 627–49.

Hogrefe, G. Jurgen, Heinz Wimmer, and Josef Perner. 1986. Ignorance vs. false belief: A developmental lag in attribution of epistemic states. *Child Development* 57: 567–82.

Hookway, Christopher. 1988. *Quine: language, experience and reality*. Stanford, CA: Stanford University Press.

Huemer, Michael. 2001. *Skepticism and the veil of perception*. Lanham, MD: Rowman and Littlefield.

Hume, David. 1975. *Enquiries concerning human understanding and concerning the principles of morals*. Eds. L. A. Selby-Bigge and P.H. Nidditch. 3rd edition. Oxford, UK: Clarendon Press.

Hume, D. 1978. *A treatise of human nature*. Ed. L.A. Selby-Bigge. 2nd ed., Oxford, UK: Clarendon Press.

Jacob, Pierre. 2002. The scope and limits of mental simulation. In *Simulation and knowledge of action*, eds. J. Dokic and J. Proust, 87–106. Amsterdam/Philadelphia: John Benjamins Publishing Co.

Jackson, Frank. 1998. *From metaphysics to ethics: a defence of conceptual analysis.* Oxford: Oxford

University Press.

Johnsen, Bredo. 2005. How to read “Epistemology naturalized”. *Journal of Philosophy* 102, 78–93.

Kelley, David. 1984. A theory of abstraction. *Cognition and Brain Theory* 7 (3&4): 329–57.

———. 1986. *The evidence of the senses: a realist theory of perception*. Baton Rouge: Louisiana State University Press.

Kim, Jaegwon. 1985. Psychophysical laws. In *Actions and events: perspectives on the philosophy of Donald Davidson*, eds. E. Lepore and B. McGlaughlin, 369–86. Oxford: Basil Blackwell.

———. 1988. What is “naturalized epistemology”? In *Philosophical Perspectives*, 2, *Epistemology*, ed. J. Tomberlin, 381–405.

———. 2005. *Physicalism, or something near enough.* Princeton, NJ: Princeton University Press.

Kirk, Robert. 2000. Indeterminacy of translation. In *The Cambridge companion to Quine*, ed. Roger Gibson, 151–80. New York: Cambridge University Press.

Kitcher, Philip. 1992. The naturalists return. *The Philosophical Review* 101 (1): 53–114.

———. 1993. *The advancement of science*. New York: Oxford University Press.

———. 2000. A priori knowledge revisited. In *New essays on the a priori*, eds. P. Boghossian and C. Peacocke, 65–91. Oxford: Oxford University Press.

Koppelberg, Dirk. 1998. Foundationalism and coherentism reconsidered. *Erkenntnis* 49: 255–83.

Kornblith, Hilary. 2002. *Knowledge and its place in nature*. Oxford, UK: Clarendon Press.

Kripke, Saul. 1972. Naming and necessity. In *The semantics of natural language*, eds. G. Harman and D. Davidson. Dordrecht, Holland: Reidel.

Kukla, Andre. 1993. Laudan, Leplin, empirical equivalence and underdetermination. *Analysis* 53: 1–7.

———. 1996. Does every theory have empirically equivalent rivals? *Erkenntnis* 44: 137–66.

Lange, Marcus. 2004. Would “direct realism” resolve the classical problem of induction? *Nous* 38 (2): 197–232.

Laudan, Larry. 1981. *Science and hypothesis: historical essays on scientific methodology*. Boston: D. Reidel Publishing Company.

———. 1984. *Science and values*. Berkeley: University of California Press.

———. 1987. Progress or rationality? The prospects for normative naturalism. *American Philosophical Quarterly* 24: 19–31.

———. 1990a Aim-less epistemology? *Studies in the History and Philosophy of Science* 21 (2): 315–22.

———. 1990b. Normative naturalism. *Philosophy of Science* 57, 44–59.

Laudan, Larry. and Jarrett Leplin. 1991. Empirical equivalence and underdetermination. *Journal of Philosophy* 88: 449–72.

———. 1993. Determination undeterred: Reply to Kukla. *Analysis* 53: 8–16.

Le Morvan, Pierre. 2004. Arguments against direct realism and how to counter them. *American Philosophical Quarterly* 41 (3): 221–34.

Leplin, Jarrett. 1982. The assessment of auxiliary hypotheses. *British Journal for the Philosophy of Science* 33: 235–249.

———. 1997. *A novel defense of scientific realism*. New York: Oxford University Press.

Leslie, Alan, A and Stephanie Keeble. 1987. Do six-month-old infants perceive causality? *Cognition* 25: 265–88.

Lewis, David. 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50 (3): 249–58.

———. 1995. Reduction of mind. In S. Gutenplan (ed.) *A companion to the philosophy of mind*. Oxford: Blackwell: 412–31.

Macherey, Edouard, Ron Mallon, Shaun Nichols, and Stephen Stich. 2004. Semantics, cross-cultural style. *Cognition* 92: B1–B12.

Maffie, James. 1990. Recent work on naturalized epistemology. *American Philosophical Quarterly* 27 (4): 281–93.

Maher, Patrick. 2004. Bayesianism and irrelevant conjunction. *Philosophy of Science* 71: 515–20.

———. 2006. Confirmation theory. In *The encyclopedia of philosophy*, ed. Donald M. Borchert, 2nd ed. New York: Macmillan.

Majors, Brad and Sarah Sawyer. 2005. The epistemological argument for content externalism. *Philosophical Perspectives* 19: 257–80.

Massimi, Michela. 2004. What demonstrative induction can do against the threat of underdetermination: Bohr, Heisenberg, and Pauli on spectroscopic anomalies (1921–24). *Synthese* 140: 243–77.

McDowell, John. 1996. *Mind and world*. Cambridge, MA: Harvard University Press.

Michotte Albert. 1963. *The perception of causality*. New York: Basic Books.

Millikan, Ruth. 1984. *Language, thought and other biological categories.* Cambridge, MA: MIT Press.

Mitchell, Peter and Riggs, Kevin. 2000. *Children's reasoning and the mind.* East Sussex, UK: Psychology Press.

Modee, Johan. 2000. Observation sentences and joint attention. *Synthese* 124: 221–38.

Moreland, J.P. 1998. Should a naturalist be a supervenient physicalist? *Metaphilosophy* 29: 35–57.

Neander, Karen. 2004. Teleological theories of mental content. In *The Stanford encyclopedia of philosophy (Summer 2004 edition)*, ed. Edward N. Zalta. http://plato.stanford.edu/archives/sum2004/entries/content-teleological (accessed July 6, 2007).

Nelson, Katherine. 1988. Constraints on word learning? *Cognitive Development* 3: 221–46.

Newton, Isaac. 1952. *Opticks*. New York: Dover Reprint.

Newton-Smith, W.H. 2001. Underdetermination of theory by data. In *A companion to the philosophy of science*, ed. W.H. Newton-Smith, 532–6. Oxford, UK: Blackwell.

Nichols, Shaun and Stephen Stich. 2003. *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds*. New York: Oxford University Press.

Nichols, Shaun, Stephen Stich, and Jonathan Weinberg. 2003. Meta-skepticism: meditations on ethno-epistemology. In *The skeptics*, ed. S. Luper, 227–47. Aldershot, UK: Ashgate Publishing.

Nimtz, Christian. 2005. Reassessing referential indeterminacy. *Erkenntnis* 62: 1–28.

Noë, Alva. 2002. On what we see. *Pacific Philosophical Quarterly* 83: 57–80.

Nolan, Daniel. 2002. Modal Fictionalism. In *The Stanford encyclopedia of philosophy (Summer 2002 edition)*, ed. Edward N. Zalta. http://plato.stanford.edu/archives/sum2002/entries/fictionalism-modal (accessed July 6, 2007).

Norton, John. 1994. Science and certainty. *Synthese* 99 (1): 3–22.

———. 2003a. A material theory of induction. *Philosophy of Science* 70: 647–70.

———. 2003b. Must evidence underdetermine theory? http://philsci-archive.pitt.edu/archive/00001257 (accessed August 4, 2006).

Nozick, Robert. 1981. *Philosophical explanations*. Cambridge, MA: Harvard University Press.

Okasha, Samir. 1997. Laudan and Leplin on empirical equivalence. *British Journal of the Philosophy of Science* 48: 2516.

———. 2003. Scepticism and its sources. *Philosophy and Phenomenological Research* 67 (3): 610–32.

Onishi, Kris and Reneé Baillergeon. 2005. Do 15-month-old infants understand false beliefs? *Science* 308: 255–8.

Orenstein, Alex. 1997. Arguing from inscrutability of reference to indeterminacy of meaning. *Revue Internationale de Philosophie* 4: 507–19.

Peijnenburg, Jeanne and Ronald Hünneman. 2001. Translations and theories: On the difference between indeterminacy and underdetermination. *Ratio* 14: 18–32.

Perner, Josef. 1991. *Understanding the representational mind*. Cambridge, MA: MIT Press.

———. 2000. About + belief + counterfactual. In *Children's reasoning and the mind*, eds. P. Mitchell and K. Riggs, 367–401. East Sussex, UK: Psychology Press.

Perner, Josef., Sarah Baker and Deborah Hutton. 1994. Prelief: the conceptual origins of belief and pretence. In *Children's early understanding of the mind: origins and development*, eds. C. Lewis and P. Mitchell, 261–86. Hove, UK: Lawrence Erlbaum Associates.

Perner, Josef and Ted Ruffman. 2005. Infants' insight into the mind: how deep? *Science* 308: 214–16.

Picardi, Eva. 2000. Empathy and charity. In *Quine: naturalized epistemology, perceptual knowledge and ontology*, eds. L. Decock and L. Horsten, 121–34. Amsterdam/Atlanta: Rodopi. Originally published in *Poznan Studies in the Philosophy of the Sciences and the Humanities* 70: 121–34.

Pollock, John. 1986. *Contemporary theories of knowledge*. London: Hutchinson.

Pollock, John and Iris Oved. 2005. Vision, knowledge and the mystery link. *Philosophical Perspectives* 19: 309–51.

Pratt, Chris and Peter Bryant. 1990. Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development* 61: 973–82.

Price, Hew. 2004. Naturalism without representation. In *Naturalism in question*, eds. M. de Caro and D. Macarthur, 71–88. Cambridge, MA: Harvard University Press.

Prinz, Jesse. 2002. *Furnishing the mind: concepts and their perceptual basis.* Cambridge, MA: MIT Press/Bradford Books.

Pritcard, Duncan. 2004. Epistemic deflationism. *The Southern Journal of Philosophy* 42: 103–34.

Putnam, Hilary. 1975. The nature of mental states. In *Philosophical papers*, vol. 2, 429–40. Cambridge: Cambridge University Press.

———. 1994. Sense, nonsense, and the senses: an inquiry into the powers of the human mind. *Journal of Philosophy* 91 (9): 445–517.

———. 1995. A comparison of something with something else. In *Words and life*, 330–50. Cambridge, MA: Harvard University Press.

———. 2004. The content and appeal of "naturalism." In *Naturalism in question*, eds. M. De Caro and D. Macarthur, D., 59–70. Cambridge, MA: Harvard University Press.

Quine, Willard Van Orman. 1947. On universals. *The Journal of Symbolic Logic* 12 (3): 74–84.

———. 1953a, Notes on the theory of reference. In *From a logical point of view*, 130–8. Cambridge, MA, Harvard University Press.

———. 1953b, Two dogmas of empiricism. In *From a logical point of view*, 20–46. Cambridge, Mass., Harvard University Press.

———. 1953c. Reference and modality. In *From a logical point of view*, 139–59. Cambridge, Mass., Harvard University Press.

———. 1960. *Word and object*. Cambridge, MA: Harvard University Press.

———. 1969a. Epistemology naturalized. In *Ontological relativity and other essays*, 69–90. New York: Columbia University Press.

———. 1969b. Natural kinds. In *Ontological relativity and other essays*, 114–38. New York: Columbia University Press.

———. 1969c. Ontological relativity, in *Ontological relativity and other essays*, 26–68. New York: Columbia University Press.

———. 1969d. Reply to Chomsky. In *Words and objections: essays on the work of W.V. Quine*, eds. D. Davidson and J. Hintikka, 302–11. Dordrecht, Holland: Reidel. Originally published in *Synthese* 19 (1–2): 274–83, 1968.

———. 1970. On the reasons for the indeterminacy of translation. *Journal of Philosophy* 67: 178–83.

———. 1974. *The roots of reference*. LaSalle, IL: Open Court Press.

———. 1975a. On empirically equivalent systems of the world. *Erkenntnis* 9: 313–28.

———. 1976. A comment on Grünbaum's claim. In *Can theories be refuted?*, ed. Sandra G. Harding, 132. Boston: D. Reidel Publishing Company.

———. 1979. Comments on Newton-Smith. *Analysis* 39(3): 66–7.

———. 1981a. Reply to Stroud. *Midwest Studies in Philosophy* 6: 473–5.

———. 1981b. Things and their place in theories. In *Theories and things*, 1–23. Cambridge, MA: Harvard University Press.

———. 1986a. Reply to Morton White. In *The philosophy of W.V. Quine,* eds. L. Hahn and P. Schilpp, 663–5. La Salle, IL: Open Court.

———. 1986b. Reply to Roger Gibson. In *The philosophy of W.V. Quine,* eds. L. Hahn and P. Schilpp. La Salle, IL: Open Court.

———. 1987. Indeterminacy of translation again. *Journal of Philosophy* 84: 5–10.

———. 1990. Comment on Bergstrom. In *Perspectives on Quine*, eds. Robert Barrett and Roger Gibson, 53–4. Oxford: Basil Blackwell.

———. 1992. *Pursuit of truth*. Cambridge, MA: Harvard University Press.

———. 1994. Promoting extensionality. *Synthese* 98 (1): 143–51.

———. 1995. *From stimulus to science*. Cambridge, MA: Harvard University Press.

———. 2004a. Confessions of a confirmed extensionalist. In *Quintessence: basic readings from the philosophy of W.V. Quine*, ed. Roger Gibson, 329–37. Cambridge, MA: Harvard University Press. Originally published in *Future pasts*, eds., J. Floyd and S. Shieh, 215–21. New York: Oxford, 2001.

———. 2004b. Naturalism; or, living within one's means. In *Quintessence: basic readings from the philosophy of W.V. Quine*, ed. Roger Gibson, 275–286. Cambridge, MA: Harvard University Press. Originally published in *Dialectica* 49 (2–4): 251–61.

———. 2004c. On simple theories of a complex world. In *Quintessence: basic readings from the philosophy of W.V. Quine*, ed. Roger Gibson, 210–13. Cambridge, MA: Harvard University Press. Originally published in *Synthese*, 15: 103–106, 1963.

———. 2004d. The nature of natural knowledge. In *Quintessence: basic readings from the philosophy of W.V. Quine*, ed. Roger Gibson, 287–300. Cambridge, MA: Harvard University Press. Originally published in *Mind and Language*, ed. S. Gutenplan, 67–81. Oxford: Oxford University Press.

———. 2004e. The scope and language of science. In *Quintessence: basic readings from the philosophy of W.V. Quine*, ed. Roger Gibson, 193–209. Cambridge, MA: Harvard University Press. Originally published in *British Journal for the Philosophy of Science* 8(29): 1–17, 1957.

Raatikainen, Panu. 2006. Is Quine a verificationist? *Southern Journal of Philosophy* 41 (3): 399–410.

Ramsey, William. 1992. Prototypes and conceptual analysis. *Topoi* 11: 59–70.

Rand, Ayn. 1990. *Introduction to Objectivist epistemology.* 2nd ed. New York: Meridian.

Rey, Georges. 1998. A naturalistic a priori*.* *Philosophical Studies* 92: 25–43.

Reynolds, Steven. 1991. Knowing how to believe with justification. *Philosophical Studies* 64: 273–92.

Riggs, Kevin, Donald Peterson, Elizabeth Robinson, and Peter Mitchell. 1998. Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development* 13: 73–90.

Romanos, George. 1983. *Quine and analytic philosophy*. Cambridge, MA: MIT Press.

Rorty, Richard. 1979. *Philosophy and the mirror of nature*. Princeton, N.J.: Princeton University Press.

Rosch, Eleanor. 1978. Principles of categorization. In *Cognition and categorization,* eds. E. Rosch and B.B. Lloyd, 27–48. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosen, Gideon. 1990. Modal fictionalism. *Mind* 99 (395): 327–54.

Rosenberg, Alexander. 1996. A field guide to recent species of naturalism. *British Journal of the Philosophy of Science* 47: 1–29.

———. 1999. Naturalistic epistemology for eliminative materialists. *Philosophy and Phenomenological Research* 59: 335–57.

Ryle, Gilbert. 1949. *The concept of mind*. New York: University Paperbacks.

Sarkar, Husain. 2000. Empirical equivalence and underdetermination. *International Studies in the Philosophy of Science* 14 (2): 187–197.

Schroeter, Laura. 2004. The limits of conceptual analysis. *Pacific Philosophical Quarterly* 85: 425–53.

Scholl, Brian, and Patrice Tremoulet. 2000. Perceptual causality and animacy. *Trends in Cognitive Science* 4 (8): 299–305.

Salmon, Wesley. 1998. *Causality and explanation*. New York: Oxford University Press.

Sellars, Wilfrid. 1963. Empiricism and the philosophy of mind. In *Science, perception and reality*, 127–96. London: Routledge.

Shatz, Marilyn, Henry Wellman, and Sharon Silber. 1983. The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition* 14: 301–21.

Siegel, Harvey. 1989. Philosophy of science naturalized? Some problems with Giere's naturalism. *Studies in the History and Philosophy of Science* 20 (3): 365–75.

———. 1990. Laudan's normative naturalism. *Studies in the History and Philosophy of Science* 21 (2): 295–313.

———. 1996. Instrumental rationality and naturalized philosophy of science. *Philosophy of Science* 63 (Proceedings): S116–24.

Sinclair, Robert. 2004. When naturalized epistemology turns normative: Kim on the failures of Quinean epistemology. *Southwest Philosophy Review* 20: 53–67.

Smart, J.J.C. 1959. Sensations and brain processes. *Philosophical Review* 68: 141–56.

Soames, Scott. 2005. *Philosophical analysis in the twentieth century, Volume 1: The dawn of analysis*, Princeton, NJ: Princeton University Press.

Sober, Elliot. 1985. Panglossian functionalism and the philosophy of mind. *Synthese* 64: 165–93.

———. 1999. Testability. *Proceedings and addresses of the American Philosophical Association* 73: 47–76.

Sosa, E. 1980. The foundations of foundationalism. *Nous* 14: 547–65.

Stanford, P. Kyle and Philip Kitcher. 2000. Refining the causal theory of reference for natural kind terms, *Philosophical Studies* 97: 99–129.

Stich, Stephen. 1983. *From folk psychology to cognitive science: the case against belief.* Cambridge, MA: MIT Press.

———. 1988. Reflective equilibrium, analytic epistemology and the problem of cognitive diversity. *Synthese* 74: 391–413.

———. 1990. *The fragmentation of reason*. Cambridge, MA: MIT Press.

———. 1992. What is a theory of mental representation? *Mind* 101: 243–61.

———. 1996. *Deconstructing the mind.* New York: Oxford University Press.

Stroud, Barry. 1981. The significance of naturalized epistemology. *Midwest Studies in Philosophy,* 6: 455–471.

———. 1984. *The significance of philosophical skepticism*. Oxford: Oxford University Press.

Turvey, Michael, Robert Shaw, Edward Reed and William Mace. 1981. Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition* 9: 237–304.

Tye, Michael. 1992. Naturalism and the mental. *Mind* 101 (403): 421–41.

Vahid, Hamid. 2004. Doubts about epistemic supervenience. *Journal of Philosophical Research* 29: 153–72.

Van Cleve, J. 1985. Epistemic supervenience and the circle of belief. *Monist* 68: 90–104.

Waskan, Jonathan. 2006. *Models and cognition: prediction and explanation in everyday life and in science*. Cambridge, MA: MIT Press.

Weiskopf, Daniel. 2005. Mental mirroring and the origin of attributions. *Mind and Language* 20 (5):

495–520.

Wellman, Henry, David Cross, and Julanne Watson. 2001. Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development* 72: 655–84.

Williams, Michael. 1996. *Unnatural doubts: epistemological realism and the basis of skepticism* Princeton, NJ: Princeton University Press.

Williamson, Timothy. 2002. *Knowledge and its limits*. New York: Oxford University Press.

Wimmer, Heinz. and Josef Perner. 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13: 103–28.

Wimmer, Heinz, G. Jurgen Hogrefe, and Josef Perner. 1988. Children's understanding of informational access as a source of knowledge. *Child Development* 59: 386–96.

Wittgenstein, Ludwig. 1969. *On certainty*, eds. G.E.M. Anscombe and G.H. von Wright. New York: Harper Torchbooks.

Wright, Larry. 1976. *Teleological explanations.* Berkeley, CA: University of California Press.

Yolton, John. 1979. As in a looking-glass: perceptual acquaintance in eighteenth-century Britain. *Journal of the History of Ideas* 40(2): 207–34.

AUTHOR'S BIOGRAPHY

Benjamin Bayer was born in Appleton, WI on July 15, 1976. He first became interested in philosophy through involvement in high school debate (specifically, the "Lincoln-Douglas" format), which introduced him to ethical and political philosophy. Originally intending to major in international relations, he attended Georgetown University's Edmund A. Walsh School of Foreign Service in the Fall of 1994. After some involvement with the Model United Nations club, Ben realized he was not very diplomatic and was more interested in debating the philosophical justification of various policies than in forming committees to compromise about them. His time spent as a research assistant to Thomas Beauchamp at the Kennedy Institute for Ethics also convinced him that he would rather make a career out of philosophy. After transferring to Lawrence University in Appleton, WI, he picked up the philosophy major and also attempted a double major in physics. His second major was downgraded to a minor when he discovered he was unaccomplished at differential equations. But this freed up more time to spend on philosophy, particularly under the useful tutelage of Tom Ryckman. But he never lost his interest in science, which is reflected in the present dissertation. Bayer graduated from Lawrence in June of 1998, and proceeded to the University of Illinois, where he received his Ph.D. in 2007.